

Source-Channel Coding Theorems for the Multiple-Access Relay Channel

Yonathan Murin[†], Ron Dabora[†], Deniz Gündüz^{*}

[†]Dept. of Electrical and Computer Engineering, Ben-Gurion University, Israel

^{*}Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain

Email: moriny@ee.bgu.ac.il, ron@ee.bgu.ac.il, deniz.gunduz@cttc.es

Abstract

Reliable transmission of arbitrarily correlated sources over multiple-access relay channels (MARC) and multiple-access broadcast relay channels (MABRC) is considered. In MARC only the destination is interested in a reconstruction of the sources, while in MABRC both the relay and the destination want to reconstruct the sources. In addition to arbitrary correlation among the source signals at the users, both the relay and the destination are assumed to have side information correlated with the source signals. Sufficient conditions for reliable communication based on operational separation, as well as necessary conditions on the achievable source-channel rates are characterized. Since operational separation is generally not optimal for MARC and MABRC, sufficient conditions for reliable communication using joint source-channel coding schemes based on a combination of the correlation preserving mapping technique with binning are also derived. For correlated sources transmitted over fading Gaussian MARC and MABRC, conditions under which informational separation is optimal are found.

Index Terms

Multiple-access relay channel; separation theorem; fading; joint source and channel coding.

I. INTRODUCTION

The multiple-access relay channel (MARC) models a network in which several users communicate with a single destination with the help of a relay [1]. Examples of such a network are sensor and ad-hoc networks in which an intermediate relay is introduced to assist the communication from the source terminals to the destination. MARC is a fundamental multi-terminal channel model that generalizes both the multiple access channel (MAC) and the relay channel models, and has received a lot of attention in the recent years [1], [2], [3], [5]. If the relay terminal also wants to decode the source messages, the model is called the multiple-access broadcast relay channel (MABRC).

While previous work on MARC considered independent sources at the terminals, here we allow arbitrary correlation among the sources that should be transmitted to the destination in a lossless fashion, and also consider

This work was partially supported by the European Commission's Marie Curie IRG Fellowship PIRG05-GA-2009-246657 under the Seventh Framework Programme.

the availability of side information correlated with the sources at the relay and the destination. The problem we address is to determine whether the given sources can be losslessly transmitted to the destination for a given number of channel uses per source sample, which is defined as the *source-channel rate*.

In [6] Shannon showed that a source can be reliably transmitted over a point-to-point channel, if and only if its entropy is less than the channel capacity. Hence, a simple comparison of the rates of the optimal source and channel codes for the respective source and channel suffices to conclude if reliable communication is feasible. This is called the *separation theorem*. The implication of the separation theorem is that independent design of source and channel codes is optimal. However, the optimality of source-channel separation does not generalize to multiuser networks [8], [9], [10], and, in general the source and channel codes need to be jointly designed for every particular source and channel combination.

The fact that the MARC generalizes both the MAC and the relay channel models reveals the difficulty of the problem studied here. The capacity of the relay channel, which corresponds to the special case of our problem with a single source terminal and no side information, remains open despite the ongoing efforts. While the capacity of the MAC is known for independent messages, the optimal joint source-channel code in the case of correlated sources is not known in general [9]. Accordingly, our goal in this paper is to identify lower and upper bounds for the achievable source-channel rate in MARC and MABRC models. We consider only decode-and-forward (DF) based achievability schemes, such that the relay terminal decodes both source signals before cooperating to forward them to the destination. Naturally, DF-based achievable schemes for MARC directly apply to the MABRC model as well. Moreover, we characterize the optimal source-channel rate in some special cases. Our contributions are listed below.

- We establish an operational separation based achievable source-channel rate for MARCs. The scheme is based on the DF scheme and uses irregular encoding, successive decoding at the relay and backward decoding at the destination. We show that for MARCs with correlated sources and side information, DF with irregular encoding yields a higher achievable source-channel rate than the rate achieved by DF with regular encoding. The achievability of this rate applies directly to the MABRCs.
- We derive necessary conditions on the achievable source-channel rate of MARCs (and MABRCs).
- We also study MARCs and MABRCs subject to independent and identically distributed (i.i.d.) fading, for both phase and Rayleigh fading. We find conditions for both the phase and Rayleigh fading models, under which informational separation is optimal. Additionally, we find conditions for the optimality of separation for fading MABRCs. *This is the first time the optimality of separation is shown for some special MARC and MABRC models.* Note that these models are not degraded in the sense of [11].
- We derive two joint source-channel DF-based achievability schemes for MARCs and MABRCs for a source-channel rate $\kappa = 1$. Both proposed schemes use a combination of binning and the correlation preserving mapping (CPM) technique proposed in [9]. While in the first scheme joint source-channel coding is used for encoding information to the relay and binning is used for encoding information to the destination; in the second scheme binning is used for encoding information to the relay and source-channel coding is used for

encoding information to the destination. A comparison between the regions achieved by the two schemes reveals a tradeoff: while the relay achievable region of the former is larger, the destination achievable region of the latter is larger. To the best of our knowledge this is the first *joint source-channel achievability scheme, based on the CPM technique, proposed for a multiuser network with a relay.*

Prior Work

The MARC has been extensively studied from a channel coding perspective. Achievable rate regions for the MARC are derived in [2], [3] and [4]. In [2] Kramer et al. derive an achievable rate region for the MARC with independent messages. The coding scheme employed in [2] is based on decode-and-forward relaying, and uses regular encoding, successive decoding at the relay and backward decoding at the destination. In [3] it is shown that, in contrast to the classic relay channel, different DF schemes yield different rate regions for the MARC. In particular, backward decoding can support a larger rate region than sliding window decoding. Another DF-based coding scheme which uses offset encoding, successive decoding at the relay and sliding-window decoding at the destination is presented in [3]. Outer bounds on the capacity region of MARCs are obtained in [4]. More recently, the capacity regions of two classes of MARCs are characterized in [5].

In [7] Shamai and Verdú considered the availability of correlated side information at the receiver in a point-to-point scenario, and showed that source-channel separation still holds. The availability of receiver side information enables transmitting the source reliably over a channel with a smaller capacity compared to the capacity needed in the absence of receiver side information. In [9] Cover et al. derived finite-letter sufficient conditions for communicating discrete, arbitrarily correlated sources over a multiple-access channel (MAC); however, these conditions were later shown by Dueck [12] to be sufficient but not necessary. The main technique used by Cover et al. is called *correlation preserving mapping (CPM)* in which the channel codewords are correlated with the source sequences, resulting in correlation between the channel inputs. CPM is extended to source coding with side information for the MAC in [13] and to broadcast channels with correlated sources in [14] (with a correction in [15]).

In [16] Tuncel distinguishes between two types of source-channel separation. *Informational separation* refers to classical separation in the Shannon sense. *Operational separation* refers to statistically independent source and channel codes that are not necessarily the optimal codes for the underlying source or the channel. Tuncel also shows that for broadcasting a single source to multiple receivers, each with its own side information, operational separation is optimal while informational separation is not.

In [10] Gündüz et al. obtain necessary and sufficient conditions for the optimality of informational separation in MACs with correlated sources and side information at the receiver. Gündüz et al. also obtain necessary and sufficient conditions for the optimality of operational separation for the compound MAC setup. Transmission of arbitrarily correlated sources over interference channels (ICs) is studied in [17], in which Salehi and Kurtas extend the CPM technique to ICs; however, when the sources are independent, the region derived in [17] does not specialize to the Han and Kobayashi (HK) region, [18]. An achievable region based on the CPM technique which specializes to the HK region is derived in [19]. The case of independent sources over ICs with correlated receiver side information

is studied in [20]. Liu et al. show that source-channel separation is optimal when each receiver has access to side information correlated with its own desired source. When each receiver has access to side information correlated with the interfering transmitter's source, Liu et al. provide sufficient conditions for reliable transmission based on a joint source-channel coding scheme using Han-Kobayashi superposition encoding and partial interference cancellation.

Lossless transmission over a relay channel with correlated side information is studied in [21], [22], [23], [25] and [26]. In [22] Gündüz and Erkip propose a DF based achievability scheme and show that operational separation is optimal for physically degraded relay channels as well as for cooperative relay-broadcast channels. This scheme is extended to multiple relay networks in [23].

Prior work on source transmission over fading channels is mostly limited to the point-to-point setup (see [24] and references therein). However, many practical sensor network scenarios involve transmission of correlated sources over fading channels. We consider two types of fading models: phase fading and Rayleigh fading. Phase fading models apply to high-speed microwave communications where the oscillator's phase noise and the system timing jitter are the key impairments. Phase fading is also the major impairment in communication systems that employ orthogonal frequency division multiplexing [27], as well as in some applications of naval communications. Additionally, phase fading can be used to model systems which use dithering to decorrelate signals [28]. Rayleigh fading models are very common in wireless communications and apply to mobile communications in the presence of multiple scatterers without line-of-sight [29]. The key similarity between the two fading models is the uniformly distributed phase of the fading process. The phase fading and Rayleigh fading models differ in the behavior of the fading magnitude component, which is fixed for the former but varies following a Rayleigh distribution for the latter. For cooperative multi-user scenarios, phase-fading models have been considered for MARCs [2], [4], [30], for broadcast-relay channels (BRCs) [2] and for interference channels [31]. Rayleigh fading models have been considered for relay channels in [32], [33] and [34], and for MARCs in [30].

The rest of this paper is organized as follows: in Section II the model and notations are presented. In Section III an operational separation-based achievable source-channel rate is presented. In Section IV necessary conditions on the achievable source-channel rates are derived. The optimality of separation for correlated sources transmitted over fading Gaussian MARCs is studied in Section V and in Section VI two joint source-channel achievable schemes are derived. Concluding remarks are provided in Section VII.

II. NOTATIONS AND MODEL

In the following we denote the set of real numbers with \Re , and we use \mathbb{C} to denote the set of complex numbers. We denote random variables with upper case letters e.g. X , Y , and their realizations with lower case letters x , y . A discrete random variable X takes values in a set \mathcal{X} . We use $|\mathcal{X}|$ to denote the cardinality of a finite, discrete set \mathcal{X} , $p_X(x)$ to denote the probability mass function (p.m.f.) of a discrete RV X on \mathcal{X} , and $f_X(x)$ to denote the probability density function (p.d.f.) of a continuous RV X on \mathbb{C} . For brevity we may omit the subscript X when it is the uppercase version of the sample symbol x . We use $p_{X|Y}(x|y)$ to denote the conditional distribution of X given Y . We denote vectors with boldface letters, e.g. \mathbf{x} , \mathbf{y} ; the i 'th element of a vector \mathbf{x} is denoted by x_i and we use

\mathbf{x}_i^j where $i < j$ to denote $(x_i, x_{i+1}, \dots, x_{j-1}, x_j)$; x^j is a short form notation for x_1^j , and unless specified otherwise $\mathbf{x} \triangleq x^n$. We denote the empty set with ϕ , and the complement of the set B by B^c . We use $H(\cdot)$ to denote the entropy of a discrete random variable and $I(\cdot; \cdot)$ to denote the mutual information between two random variables, as defined in [35, ch. 2, ch. 9]. We use $A_\epsilon^{*(n)}(X)$ to denote the set of ϵ -strongly typical sequences w.r.t. distribution $p_X(x)$ on \mathcal{X} , as defined in [36, ch. 5.1], and $A_\epsilon^{(n)}(X)$ to denote the set of ϵ -weakly typical sequences as defined in [35, ch. 3]. When referring to a typical set we may omit the random variables from the notation, when these variables are clear from the context. We use $\mathcal{CN}(a, \sigma^2)$ to denote a proper, circularly symmetric, complex Normal distribution with mean a and variance σ^2 [37], and $\mathbb{E}\{\cdot\}$ to denote stochastic expectation. We use $X - Y - Z$ to denote a Markov chain formed by the random variables X, Y, Z as defined in [35, ch. 2.8].

A. Problem Formulation for Lossless Source Coding for MARCs and MABRCs

The MARC consists of two transmitters (sources), a receiver (destination) and a relay. Transmitter i has access to the source sequence S_i^m , for $i = 1, 2$. The receiver is interested in the lossless reconstruction of the source sequences observed by the two transmitters. The relay has access to side information W_3^m and the receiver has access to side information W^m . The objective of the relay is to help the receiver decode the source sequences. It is also assumed that the side information at the relay and the receiver are correlated with the source sequences. For the MABRC the relay is also interested in a lossless reconstruction of the source sequences. Figure 1 depicts the MABRC with side information setup. The MARC is obtained when the reconstruction at the relay is omitted.

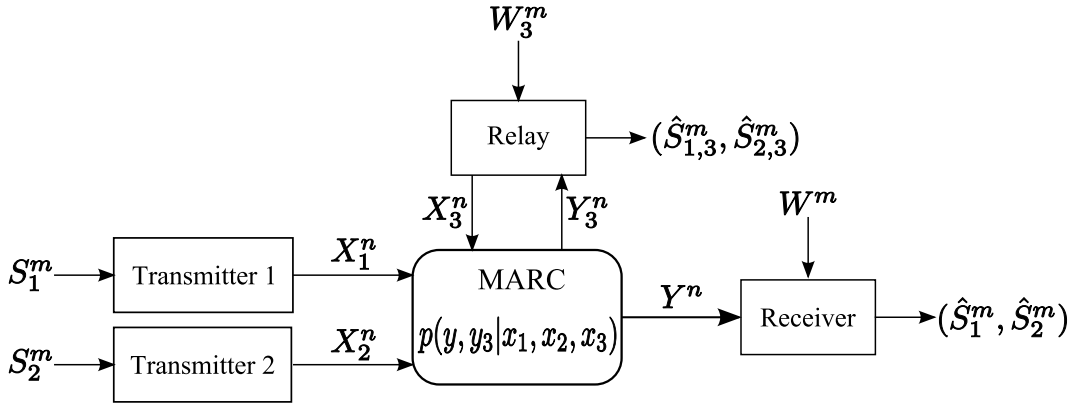


Fig. 1. Multiple-access broadcast relay channel with correlated side information. $(\hat{S}_{1,3}^m, \hat{S}_{2,3}^m)$ are the reconstructions of $(S_{1,3}^m, S_{2,3}^m)$ at the relay, and $(\hat{S}_1^m, \hat{S}_2^m)$ are the reconstructions at the destination.

The sources and the side information sequences, $\{S_{1,k}, S_{2,k}, W_k, W_{3,k}\}_{k=1}^m$, are arbitrarily correlated according to a joint distribution $p(s_1, s_2, w, w_3)$ over a finite alphabet $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{W} \times \mathcal{W}_3$, and independent across different sample indices k . All nodes know this joint distribution.

For transmission, a discrete memoryless MARC with inputs X_1, X_2, X_3 over finite input alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$, and outputs Y, Y_3 over finite output alphabets $\mathcal{Y}, \mathcal{Y}_3$, is available. The MARC is memoryless in the sense

$$p(y_k, y_{3,k} | y_1^{k-1}, y_{3,1}^{k-1}, x_1^k, x_2^k, x_3^k) = p(y_k, y_{3,k} | x_{1,k}, x_{2,k}, x_{3,k}). \quad (1)$$

Definition 1. An (m, n) source-channel code for the MABRC with correlated side information consists of two encoding functions,

$$f_i^{(m,n)} : \mathcal{S}_i^m \mapsto \mathcal{X}_i^n, \quad i = 1, 2, \quad (2)$$

a set of causal encoding functions at the relay, $\{f_{3,k}^{(m,n)}\}_{k=1}^n$, such that

$$x_{3,k} = f_{3,k}^{(m,n)}(y_{3,1}^{k-1}, w_3^m), \quad 1 \leq k \leq n, \quad (3)$$

and two decoding functions

$$g^{(m,n)} : \mathcal{Y}^n \times \mathcal{W}^m \mapsto \mathcal{S}_1^m \times \mathcal{S}_2^m, \quad (4a)$$

$$g_3^{(m,n)} : \mathcal{Y}_3^n \times \mathcal{W}_3^m \mapsto \mathcal{S}_1^m \times \mathcal{S}_2^m. \quad (4b)$$

An (m, n) code for the MARC scenario is defined as in Definition 1 with the exception that the decoding function $g_3^{(m,n)}$ does not exist.

Definition 2. Let $\hat{S}_i^m, i = 1, 2$, denote the reconstruction of $S_i^m, i = 1, 2$, respectively at the receiver. Let $\hat{S}_{i,3}^m, i = 1, 2$, denote the reconstruction of $S_i^m, i = 1, 2$, respectively at the relay. The average probability of error, $P_e^{(m,n)}$, of an (m, n) code for the MABRC is defined as

$$P_e^{(m,n)} \triangleq \Pr \left\{ \{(\hat{S}_1^m, \hat{S}_2^m) \neq (S_1^m, S_2^m)\} \cup \{(\hat{S}_{1,3}^m, \hat{S}_{2,3}^m) \neq (S_1^m, S_2^m)\} \right\}, \quad (5)$$

while for the MARC the average probability of error is defined as

$$P_e^{(m,n)} \triangleq \Pr \left\{ (\hat{S}_1^m, \hat{S}_2^m) \neq (S_1^m, S_2^m) \right\}. \quad (6)$$

Definition 3. A source-channel rate κ is said to be achievable for the MABRC if, for every $\epsilon > 0$, there exist positive integers n_0, m_0 such that for all $n > n_0, m > m_0, n/m = \kappa$ there exists an (m, n) code for which $P_e^{(m,n)} < \epsilon$. The same definition applies to the MARC.

B. Fading Gaussian MARCs

The fading Gaussian MARC is depicted in Figure 2. In fading Gaussian MARCs, the received signals at time k at the receiver and the relay are given by

$$Y_k = H_{11,k}X_{1,k} + H_{21,k}X_{2,k} + H_{31,k}X_{3,k} + Z_k \quad (7a)$$

$$Y_{3,k} = H_{13,k}X_{1,k} + H_{23,k}X_{2,k} + Z_{3,k}, \quad (7b)$$

$k = 1, \dots, n$, where Z and Z_3 are independent of each other, i.i.d., circularly symmetric, complex Normal RVs, $\mathcal{CN}(0, 1)$. The channel input signals are subject to per-symbol average power constraints: $\mathbb{E}\{|X_i|^2\} \leq P_i, i = 1, 2, 3$. In the following it is assumed that the destination knows the instantaneous channel coefficients from Transmitter $i, i = 1, 2$, and from the relay to itself, and the relay knows the instantaneous channel coefficients from both transmitters to itself. This is referred to as receiver channel state information (Rx-CSI). Note that the destination receiver does not have CSI on the links arriving at the relay, and that the relay does not have CSI on the links

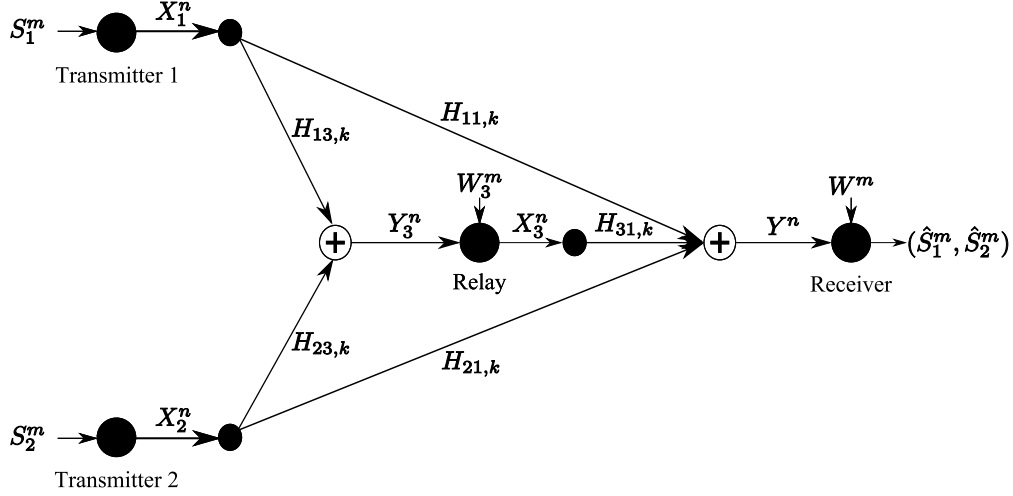


Fig. 2. Sources transmitted over fading Gaussian MARC with side information at the relay and destination (additive noises are not depicted).

arriving at the receiver. It is also assumed that the sources and the relay do not know the channel coefficients on their outgoing links (no transmitter CSI). We represent the CSI at the destination with $\tilde{H}_1 = (H_{11}, H_{21}, H_{31})$, and the CSI at the relay with $\tilde{H}_3 = (H_{13}, H_{23})$. We let $\mathfrak{H}_1 = \mathfrak{C}^3$ and $\mathfrak{H}_3 = \mathfrak{C}^2$ be the corresponding domains for the channel state vectors, and define $\tilde{H} \triangleq \{H_{li} : (l, i) \in \{(1, 1), (1, 3), (2, 1), (2, 3), (3, 1)\}\}$.

We consider two types of fading: phase fading and Rayleigh fading, and define the statistical model for each fading type as follows:

- **Phase fading channels:** The channel coefficients are given by $H_{li,k} = a_{li}e^{j\Theta_{li,k}}$, $a_{li} \in \mathfrak{R}$ are constants representing the attenuation and $\Theta_{li,k}$ are uniformly distributed over $[0, 2\pi)$, i.i.d., and independent of each other and of the additive noises Z_3 and Z . Due to Rx-CSI we can set $a_{11} = 1$.
- **Rayleigh fading channels:** The channel coefficients are given by $H_{li,k} = a_{li}U_{li,k}$, $a_{li} \in \mathfrak{R}$ are constants representing the attenuation, and $U_{li,k}$ are circularly symmetric, complex Normal RVs, $U_{li,k} \sim \mathcal{CN}(0, 1)$, i.i.d., and independent of each other and of the additive noises Z_3 and Z . We define $\tilde{U} = \{U_{li} : (l, i) \in \{(1, 1), (1, 3), (2, 1), (2, 3), (3, 1)\}\}$. Due to Rx-CSI we can set $a_{11} = 1$.

In both models the values of a_{li} are fixed and known at all users. Observe that the magnitude of the phase-fading process is constant, $|H_{li,k}| = a_{li}$, but for Rayleigh fading the fading magnitude varies between different time instances.

III. OPERATIONAL SEPARATION-BASED ACHIEVABLE RATE FOR DISCRETE MEMORYLESS MARCS AND MABRCs

In this section we derive a separation-based achievable rate for discrete memoryless (DM) MARCs and MABRCs with correlated sources and side information. The achievability is established by using Slepian-Wolf source coding [38], and a channel coding scheme similar to the one detailed in [3, Sections II, III], which employs DF with

irregular block Markov encoding, successive decoding at the relay and backward decoding at the destination.

Theorem 1. For DM MARCs and DM MABRCs with relay and receiver side information as defined in Section II-A, rate κ is achievable if,

$$H(S_1|S_2, W_3) < \kappa I(X_1; Y_3|X_2, V_1, X_3) \quad (8a)$$

$$H(S_2|S_1, W_3) < \kappa I(X_2; Y_3|X_1, V_2, X_3) \quad (8b)$$

$$H(S_1, S_2|W_3) < \kappa I(X_1, X_2; Y_3|V_1, V_2, X_3) \quad (8c)$$

$$H(S_1|S_2, W) < \kappa I(X_1, X_3; Y|X_2, V_2) \quad (8d)$$

$$H(S_2|S_1, W) < \kappa I(X_2, X_3; Y|X_1, V_1) \quad (8e)$$

$$H(S_1, S_2|W) < \kappa I(X_1, X_2, X_3; Y), \quad (8f)$$

for an input distribution that factors as

$$p(s_1, s_2, w_3, w)p(v_1)p(x_1|v_1)p(v_2)p(x_2|v_2)p(x_3|v_1, v_2). \quad (9)$$

Proof: The proof is given in Subsection III-B. ■

A. Discussion

In Figure 3 we illustrate the Markov chains for the joint distribution considered in Thm. 1.

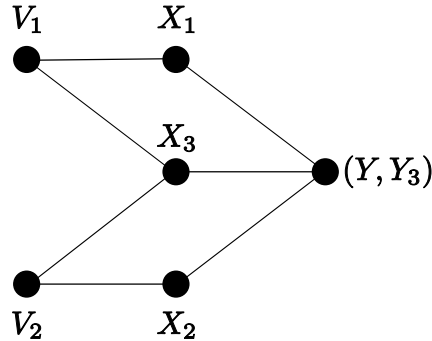


Fig. 3. Diagram of the Markov chain for the joint distribution considered in (9).

Remark 1. In Thm. 1, bounds (8a)–(8c) are constraints for decoding at the relay, while bounds (8d)–(8f) are decoding constraints at the destination.

Remark 2. The source-channel achievable rate of Thm. 1 is established by using two different Slepian-Wolf coding schemes: one for the relay and one for the destination. This requires an irregular encoding scheme for the channel code. In regular encoding, the codebooks at the source and the relay have the same size, see for example [3]. Applying regular encoding to MABRCs with correlated sources and side information leads to merging some of the constraints in (8). In particular (8a) and (8d) will be combined into the constraint

$$\max \{H(S_1|S_2, W_3), H(S_1|S_2, W)\} < \kappa \min \{I(X_1; Y_3|X_2, V_1, X_3), I(X_1, X_3; Y|X_2, V_2)\}.$$

In irregular encoding, the transmission rates to the relay and to the destination can be different, exploiting different quality of the side information. We conclude that for MABRCs with correlated sources and side information, irregular encoding yields a higher source-channel achievable rate than the one achieved by regular encoding. When the relay and destination have the same side information ($W = W_3$) then the irregular and regular schemes obtain the same achievable source-channel rates.

We note that when using *regular encoding* for MARCs, there is a single Slepian-Wolf code, hence, in the scheme used in Thm. 1 it is not required to recover the source sequences at the relay and the right-hand side (RHS) of the constraints (8a)–(8f) can be combined. For example, (8a) and (8d) will be combined into the constraint

$$H(S_1|S_2, W) < \kappa \min \{I(X_1; Y_3|X_2, V_1, X_3), I(X_1, X_3; Y|X_2, V_2)\},$$

which shows that regular encoding is more restrictive for the MARC as well.

B. Proof of Theorem 1

Fix a distribution $p(v_1)p(x_1|v_1)p(v_2)p(x_2|v_2)p(x_3|v_1, v_2)$.

1) *Code construction*: For $i = 1, 2$, assign every $\mathbf{s}_i \in \mathcal{S}_i^m$ to one of $2^{mR_i^r}$ bins independently according to a uniform distribution on $\mathcal{U}_i^r \triangleq \{1, 2, \dots, 2^{mR_i^r}\}$. We refer to these two bin sets as the relay bins. Denote these assignments by f_i^r . Independent from the relay bins, for $i = 1, 2$, assign every $\mathbf{s}_i \in \mathcal{S}_i^m$ to one of $2^{mR_i^d}$ bins independently according to a uniform distribution on $\mathcal{U}_i^d \triangleq \{1, 2, \dots, 2^{mR_i^d}\}$. We refer to these two bin sets as the destination bins. Denote these assignments by f_i^d .

For the channel codebook, for each $i = 1, 2$, let $\hat{R}_i^d = \frac{1}{\kappa} R_i^d$ and generate $2^{n\hat{R}_i^d}$ codewords $\mathbf{v}_i(u_i^d), u_i^d \in \mathcal{U}_i^d$, by choosing the letters $v_{i,k}(u_i^d), k = 1, 2, \dots, n$, independently with distribution $p_{V_i}(v_{i,k})$. Let $\hat{R}_i^r = \frac{1}{\kappa} R_i^r$, for every $\mathbf{v}_i(u_i^d)$ generate $2^{n\hat{R}_i^r}$ codewords $\mathbf{x}_i(u_i^r, u_i^d), u_i^r \in \mathcal{U}_i^r$, by choosing the letters $x_{i,k}(u_i^r, u_i^d)$ independently with distribution $p_{X_i|V_i}(x_{i,k}|v_{i,k}(u_i^d))$ for all $1 \leq k \leq n$. Finally, generate one length- n relay codeword $\mathbf{x}_3(u_1^d, u_2^d)$ for each pair $(u_1^d, u_2^d) \in \mathcal{U}_1^d \times \mathcal{U}_2^d$ by choosing $x_{3,k}(u_1^d, u_2^d)$ independently with distribution $p_{X_3|V_1, V_2}(x_{3,k}|v_{1,k}(u_1^d), v_{2,k}(u_2^d))$ for all $1 \leq k \leq n$.

2) *Encoding*: (See Table I). Consider source sequences of length Bm , $\mathbf{s}_i^{Bm} \in \mathcal{S}_i^{Bm}, i = 1, 2$. Partition each sequence into B length- m subsequences, $\mathbf{s}_{i,b}, b = 1, 2, \dots, B$. Similarly partition the corresponding side information sequences w_3^{Bm} and w^{Bm} into B length- m subsequences, $\mathbf{w}_{3,b}, \mathbf{w}_b, b = 1, 2, \dots, B$. We transmit a total of Bm source samples over $B + 1$ blocks of n channel uses each. For any fixed (m, n) with $n = \kappa m$, we can achieve a rate arbitrarily close to $\kappa = n/m$ by increasing B . i.e., $(B + 1)n/Bm \rightarrow n/m = \kappa$ as $B \rightarrow \infty$.

In block 1, transmitter $i, i = 1, 2$ observes $\mathbf{s}_{i,1}$ and finds its corresponding relay bin index $u_{i,1}^r \in \mathcal{U}_i^r$. It transmits the channel codeword $\mathbf{x}_i(u_{i,1}^r, 1)$. In block $b, b = 2, \dots, B$, source terminal i transmits the channel codeword $\mathbf{x}_i(u_{i,b}^r, u_{i,b-1}^d)$ where $u_{i,b}^r \in \mathcal{U}_i^r$ is the relay bin index for source vector $\mathbf{s}_{i,b}$, and $u_{i,b-1}^d \in \mathcal{U}_i^d$ is the destination bin index for the source vector $\mathbf{s}_{i,b-1}$. In block $B + 1$, the source terminal transmits $\mathbf{x}_i(1, u_{i,B}^d)$.

At block $b = 1$, the relay simply transmits $\mathbf{x}_3(1, 1)$. Assume that at block $b, b = 2, \dots, B, B + 1$, the relay estimates $(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$. It then finds the corresponding destination bin indices $\hat{u}_{i,b-1}^d \in \mathcal{U}_i^d, i = 1, 2$, and transmits the channel codeword $\mathbf{x}_3(\hat{u}_{1,b-1}^d, \hat{u}_{2,b-1}^d)$.

| Node | Block 1 | Block 2 | Block B | Block $B + 1$ |
|--------|---------------------------------------|--|--|--|
| User 1 | $\mathbf{s}_{1,1}$ | $\mathbf{s}_{1,2}$ | $\mathbf{s}_{1,B}$ | — |
| | $f_1^r(\mathbf{s}_{1,1}) = u_{1,1}^r$ | $f_1^r(\mathbf{s}_{1,2}) = u_{1,2}^r$ | $f_1^r(\mathbf{s}_{1,B}) = u_{1,B}^r$ | — |
| | — | $f_1^d(\mathbf{s}_{1,1}) = u_{1,1}^d$ | $f_1^d(\mathbf{s}_{1,B-1}) = u_{1,B-1}^d$ | $f_1^d(\mathbf{s}_{1,B}) = u_{1,B}^d$ |
| | $\mathbf{x}_1(u_{1,1}^r, 1)$ | $\mathbf{x}_1(u_{1,2}^r, u_{1,1}^d)$ | $\mathbf{x}_1(u_{1,B}^r, u_{1,B-1}^d)$ | $\mathbf{x}_1(1, u_{1,B}^d)$ |
| | $\mathbf{v}_1(1)$ | $\mathbf{v}_1(u_{1,1}^d)$ | $\mathbf{v}_1(u_{1,B-1}^d)$ | $\mathbf{v}_1(u_{1,B}^d)$ |
| User 2 | $\mathbf{s}_{2,1}$ | $\mathbf{s}_{2,2}$ | $\mathbf{s}_{2,B}$ | — |
| | $f_2^r(\mathbf{s}_{2,1}) = u_{2,1}^r$ | $f_2^r(\mathbf{s}_{2,2}) = u_{2,2}^r$ | $f_2^r(\mathbf{s}_{2,B}) = u_{2,B}^r$ | — |
| | — | $f_2^d(\mathbf{s}_{2,1}) = u_{2,1}^d$ | $f_2^d(\mathbf{s}_{2,B-1}) = u_{2,B-1}^d$ | $f_2^d(\mathbf{s}_{2,B}) = u_{2,B}^d$ |
| | $\mathbf{x}_2(u_{2,1}^r, 1)$ | $\mathbf{x}_2(u_{2,2}^r, u_{2,1}^d)$ | $\mathbf{x}_2(u_{2,B}^r, u_{2,B-1}^d)$ | $\mathbf{x}_2(1, u_{2,B}^d)$ |
| | $\mathbf{v}_2(1)$ | $\mathbf{v}_2(u_{2,1}^d)$ | $\mathbf{v}_2(u_{2,B-1}^d)$ | $\mathbf{v}_2(u_{2,B}^d)$ |
| Relay | $\mathbf{x}_3(1, 1)$ | $\mathbf{x}_3(\hat{u}_{1,1}^d, \hat{u}_{2,1}^d)$ | $\mathbf{x}_3(\hat{u}_{1,B-1}^d, \hat{u}_{2,B-1}^d)$ | $\mathbf{x}_3(\hat{u}_{1,B}^d, \hat{u}_{2,B}^d)$ |

TABLE I

ENCODING PROCESS FOR THE SEPARATION BASED DF SCHEME.

3) *Decoding*: The relay decodes the source sequences sequentially trying to reconstruct source blocks $\mathbf{s}_{i,b}$, $i = 1, 2$, at the end of channel block b as follows: assume that the relay knows $(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$ at the end of block $b - 1$. Hence, it can find the destination bin indices $(u_{1,b-1}^d, u_{2,b-1}^d)$. Using this information and its received signal $\mathbf{y}_{3,b}$, the relay channel decoder at time b decodes $(u_{1,b}^r, u_{2,b}^r)$, i.e., the relay bin indices corresponding to $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$ by looking for a unique pair $(\tilde{u}_1^r, \tilde{u}_2^r)$ such that:

$$\begin{aligned}
 &(\mathbf{x}_1(\tilde{u}_1^r, u_{1,b-1}^d), \mathbf{x}_2(\tilde{u}_2^r, u_{2,b-1}^d), \mathbf{v}_1(u_{1,b-1}^d), \mathbf{v}_2(u_{2,b-1}^d), \mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d), \mathbf{y}_{3,b}) \\
 &\in A_\epsilon^{*(n)}(X_1, X_2, V_1, V_2, X_3, Y_3).
 \end{aligned} \tag{10}$$

Using the relay decoded bin indices and the side information $\mathbf{w}_{3,b}$, the relay source decoder estimates $\mathbf{s}_{1,b}, \mathbf{s}_{2,b}$. More precisely, given the relay bin indices $\tilde{u}_1^r, \tilde{u}_2^r$, the relay source decoder declares $(\tilde{\mathbf{s}}_{1,b}^r, \tilde{\mathbf{s}}_{2,b}^r) \in \mathcal{S}_1^m \times \mathcal{S}_2^m$ as the decoded sequences if it is the unique pair of sequences that satisfies $f_1^r(\tilde{\mathbf{s}}_{1,b}^r) = \tilde{u}_1^r, f_2^r(\tilde{\mathbf{s}}_{2,b}^r) = \tilde{u}_2^r$ and $(\tilde{\mathbf{s}}_{1,b}^r, \tilde{\mathbf{s}}_{2,b}^r, \mathbf{w}_{3,b}) \in A_\epsilon^{*(m)}(S_1, S_2, W_3)$.

Decoding at the destination is done using backward decoding. The destination node waits until the end of channel block $B+1$ is received. It first tries to decode $(\mathbf{s}_{1,B}, \mathbf{s}_{2,B})$ using the received signal at channel block $B+1$ and its side information \mathbf{w}_B . Going backwards from the last channel block to the first, we assume that the destination knows $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ and consider decoding of $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$. From $\mathbf{s}_{i,b+1}$, $i = 1, 2$, the destination knows the relay bin indices $u_{i,b+1}^r$. At block $b+1$ the destination channel decoder first estimates the destination bin indices $(u_{1,b}^d, u_{2,b}^d)$, corresponding to $\mathbf{s}_{i,b}$ based on its received signal \mathbf{y}_{b+1} . More precisely, the destination channel decoder looks for a unique pair $(\tilde{u}_1^d, \tilde{u}_2^d)$ such that:

$$\begin{aligned}
 &(\mathbf{x}_1(u_{1,b+1}^r, \tilde{u}_1^d), \mathbf{x}_2(u_{2,b+1}^r, \tilde{u}_2^d), \mathbf{v}_1(\tilde{u}_1^d), \mathbf{v}_2(\tilde{u}_2^d), \mathbf{x}_3(\tilde{u}_1^d, \tilde{u}_2^d), \mathbf{y}_{b+1}) \\
 &\in A_\epsilon^{*(n)}(X_1, X_2, V_1, V_2, X_3, Y).
 \end{aligned} \tag{11}$$

From the decoded destination bin indices and the side information \mathbf{w}_b , the destination source decoder estimates $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$. More precisely, given the destination bin indices $(\tilde{u}_1^d, \tilde{u}_2^d)$, the destination source decoder declares $(\tilde{\mathbf{s}}_{1,b}^d, \tilde{\mathbf{s}}_{2,b}^d) \in \mathcal{S}_1^m \times \mathcal{S}_2^m$ as the decoded sequences if it is the unique pair of sequences that satisfies $f_1^d(\tilde{\mathbf{s}}_{1,b}^d) = \tilde{u}_1^d$, $f_2^d(\tilde{\mathbf{s}}_{2,b}^d) = \tilde{u}_2^d$ and $(\tilde{\mathbf{s}}_{1,b}^d, \tilde{\mathbf{s}}_{2,b}^d, \mathbf{w}_b) \in A_\epsilon^{*(m)}(S_1, S_2, W)$.

4) *Error probability analysis*: The error probability analysis is given in Appendix A.

IV. NECESSARY CONDITIONS ON THE ACHIEVABLE SOURCE-CHANNEL RATE FOR DISCRETE MEMORYLESS MARCS AND MABRCs

In this section we derive upper bounds on the achievable source-channel rate for MARCs and for MABRCs with correlated sources and side information at the relay and at the destination.

Theorem 2. Consider the transmission of arbitrarily correlated sources S_1 and S_2 over the DM MARC with relay side information W_3 and receiver side information W . Any achievable source-channel rate κ must satisfy the following constraints:

$$H(S_1|S_2, W) < \kappa I(X_1, X_3; Y|X_2) \quad (12a)$$

$$H(S_2|S_1, W) < \kappa I(X_2, X_3; Y|X_1) \quad (12b)$$

$$H(S_1, S_2|W) < \kappa I(X_1, X_2, X_3; Y), \quad (12c)$$

for some input distribution $p(x_1, x_2, x_3)$, and the constraints

$$H(S_1|S_2, W, W_3) < \kappa I(X_1; Y, Y_3|X_2, V) \quad (13a)$$

$$H(S_2|S_1, W, W_3) < \kappa I(X_2; Y, Y_3|X_1, V) \quad (13b)$$

$$H(S_1, S_2|W, W_3) < \kappa I(X_1, X_2; Y, Y_3|V), \quad (13c)$$

for some input distribution $p(v)(x_1, x_2|v)p(x_3|v)$, with $|\mathcal{V}| \leq 4$.

Proof: The proof is given in Subsection IV-A. ■

Remark 3. The RHS expressions of the constraints in (13) are similar to the broadcast bound¹ when assuming that all relay information is available at the destination.

Remark 4. Setting $\mathcal{X}_2 = S_2 = \phi$, constraints in (12) specializes to the converse of [22, Thm. 3.1] for the relay channel.

Theorem 3. Consider the transmission of arbitrarily correlated sources S_1 and S_2 over the DM MABRC with relay side information W_3 and receiver side information W . Any achievable source-channel rate κ must satisfy constraints

¹Here we use the common terminology for the classic relay channel, where in the cut-set bound the cut $I(X, X_1; Y)$ is referred to as the MAC bound while $I(X; Y, Y_1|X_1)$ is called the Broadcast bound [39, Ch. 17].

(12) as well as the following constraints:

$$H(S_1|S_2, W_3) < \kappa I(X_1; Y_3|X_2, X_3) \quad (14a)$$

$$H(S_2|S_1, W_3) < \kappa I(X_2; Y_3|X_1, X_3) \quad (14b)$$

$$H(S_1, S_2|W_3) < \kappa I(X_1, X_2; Y_3|X_3), \quad (14c)$$

for some input distribution $p(x_1, x_2, x_3)$.

Proof: The proof is given in Appendix B. ■

Remark 5. In the upper bounds detailed in constraints (12) and (14) the input distribution is of the form $p(x_1, x_2, x_3)$, while the input distribution of the bound detailed in constraints (13) is of the form $p(v)p(x_1, x_2|v)p(x_3|v)$.

A. Proofs of Thm. 2

Let $P_e^{(m,n)} \rightarrow 0$ as $n, m \rightarrow \infty$, for a sequence of encoders and decoders $f_1^{(m,n)}, f_2^{(m,n)}, f_3^{(m,n)}, g^{(m,n)}$, such that $\kappa = n/m$ is fixed. We will use Fano's inequality which states

$$\begin{aligned} H(S_1^m, S_2^m | \hat{S}_1^m, \hat{S}_2^m) &\leq 1 + mP_e^{(m,n)} \log |\mathcal{S}_1 \times \mathcal{S}_2| \\ &\triangleq m\delta(P_e^{(m,n)}), \end{aligned} \quad (15)$$

where $\delta(x)$ is a non-negative function that approaches $\frac{1}{m}$ as $x \rightarrow 0$. We also obtain

$$H(S_1^m, S_2^m | \hat{S}_1^m, \hat{S}_2^m) \geq H(S_1^m, S_2^m | Y^n, W^m, \hat{S}_1^m, \hat{S}_2^m) \quad (16a)$$

$$= H(S_1^m, S_2^m | Y^n, W^m) \quad (16b)$$

$$\geq H(S_1^m | Y^n, W^m, S_2^m), \quad (16c)$$

where (16a) follows from the fact that conditioning reduces entropy [35, Thm. 2.6.5]; (16b) follows from the fact that $(\hat{S}_1^m, \hat{S}_2^m)$ is a function of (Y^n, W^m) and (16c) follows from non-negativity of the entropy function for discrete sources.

1) *Proof of constraints (12):* Constraint (12a) is a consequence of the following chain of inequalities:

$$\begin{aligned} \sum_{k=1}^n I(X_{1,k}, X_{3,k}; Y_k | X_{2,k}) &= \sum_{k=1}^n \left[H(Y_k | X_{2,k}) - H(Y_k | X_{1,k}, X_{2,k}, X_{3,k}) \right] \\ &= \sum_{k=1}^n \left[H(Y_k | X_{2,k}) - H(Y_k | X_{1,1}^k, X_{2,1}^k, X_{3,1}^k, Y_{3,1}^{k-1}, Y^{k-1}) \right] \end{aligned} \quad (17a)$$

$$\begin{aligned} &= \sum_{k=1}^n \left[H(Y_k | X_{2,k}) \right. \\ &\quad \left. - H(Y_k | S_1^m, S_2^m, W_3^m, W^m, X_{1,1}^k, X_{2,1}^k, X_{3,1}^k, Y_{3,1}^{k-1}, Y^{k-1}) \right] \end{aligned} \quad (17b)$$

$$\geq \sum_{k=1}^n \left[H(Y_k | S_2^m, W^m, Y^{k-1}, X_{2,k}) - H(Y_k | S_1^m, S_2^m, W_3^m, W^m, Y^{k-1}) \right] \quad (17c)$$

$$= \sum_{k=1}^n \left[H(Y_k | S_2^m, W^m, Y^{k-1}) - H(Y_k | S_1^m, S_2^m, W_3^m, W^m, Y^{k-1}) \right] \quad (17d)$$

$$\begin{aligned} &= \sum_{k=1}^n I(S_1^m, W_3^m; Y_k | Y^{k-1}, S_2^m, W^m) \\ &= I(S_1^m, W_3^m; Y^n | S_2^m, W^m) \end{aligned} \quad (17e)$$

$$\geq I(S_1^m; Y^n | S_2^m, W^m) \quad (17f)$$

$$\begin{aligned} &= H(S_1^m | S_2^m, W^m) - H(S_1^m | Y^n, S_2^m, W^m) \\ &\geq mH(S_1 | S_2, W) - m\delta(P_e^{(m,n)}), \end{aligned} \quad (17g)$$

where (17a) follows from the memoryless channel assumption (see (1)); (17b) follows from the Markov relation $(S_1^m, S_2^m, W_3^m, W^m) - (X_{1,1}^k, X_{2,1}^k, X_{3,1}^k, Y_{3,1}^{k-1}, Y^{k-1}) - Y_k$ (see [40]); (17c) follows from the fact that conditioning reduces entropy; (17d) follows from the fact that $X_{2,k}$ is a deterministic function of S_2^m ; (17e) follows from the chain rule for mutual information; (17f) follows from the non-negativity of the mutual information; and (17g) follows from the memoryless sources and side information assumption and from equations (15)–(16).

Following arguments similar to those that led to (17g) we can also show

$$\sum_{k=1}^n I(X_{2,k}, X_{3,k}; Y_k | X_{1,k}) \geq mH(S_2 | S_1, W) - m\delta(P_e^{(m,n)}) \quad (18a)$$

$$\sum_{k=1}^n I(X_{1,k}, X_{2,k}, X_{3,k}; Y_k) \geq mH(S_1, S_2 | W) - m\delta(P_e^{(m,n)}). \quad (18b)$$

Using the concavity of mutual information over the set of all joint distributions $p(x_1, x_2, x_3)$, taking the limit as $m, n \rightarrow \infty$ and letting $P_e^{(m,n)} \rightarrow 0$, (17g), (18a) and (18b) result in the constraints in (12).

2) *Proof of constraints (13):* We begin by defining the following auxiliary random variable:

$$V_k \triangleq (Y_{3,1}^{k-1}, W_3^m), \quad k = 1, 2, \dots, n. \quad (19)$$

Constraint (13a) is a consequence of the following chain of inequalities:

$$\sum_{k=1}^n I(X_{1,k}; Y_k, Y_{3,k} | X_{2,k}, V_k) = \sum_{k=1}^n \left[H(Y_k, Y_{3,k} | X_{2,k}, Y_{3,1}^{k-1}, W_3^m) - H(Y_k, Y_{3,k} | X_{1,k}, X_{2,k}, Y_{3,1}^{k-1}, W_3^m) \right] \quad (20a)$$

$$= \sum_{k=1}^n \left[H(Y_k, Y_{3,k} | X_{2,k}, Y_{3,1}^{k-1}, W_3^m) - H(Y_k, Y_{3,k} | X_1^k, X_2^k, X_3^k, Y_{3,1}^{k-1}, Y^{k-1}, W_3^m) \right] \quad (20b)$$

$$\geq \sum_{k=1}^n \left[H(Y_k, Y_{3,k} | X_{2,k}, Y_{3,1}^{k-1}, Y^{k-1}, W_3^m, W^m, S_2^m) - H(Y_k, Y_{3,k} | X_1^k, X_2^k, X_3^k, Y_{3,1}^{k-1}, Y^{k-1}, W_3^m) \right] \quad (20c)$$

$$= \sum_{k=1}^n \left[H(Y_k, Y_{3,k} | X_{2,k}, Y_{3,1}^{k-1}, Y^{k-1}, W_3^m, W^m, S_2^m) - H(Y_k, Y_{3,k} | X_1^k, X_2^k, X_3^k, Y_{3,1}^{k-1}, Y^{k-1}, W_3^m, W^m, S_1^m, S_2^m) \right] \quad (20d)$$

$$= \sum_{k=1}^n \left[H(Y_k, Y_{3,k} | Y_{3,1}^{k-1}, Y^{k-1}, W_3^m, W^m, S_2^m) - H(Y_k, Y_{3,k} | X_1^k, X_2^k, X_3^k, Y_{3,1}^{k-1}, Y^{k-1}, W_3^m, W^m, S_1^m, S_2^m) \right] \quad (20e)$$

$$\geq \sum_{k=1}^n \left[H(Y_k, Y_{3,k} | Y_{3,1}^{k-1}, Y^{k-1}, W_3^m, W^m, S_2^m) - H(Y_k, Y_{3,k} | Y_{3,1}^{k-1}, Y^{k-1}, W_3^m, W^m, S_1^m, S_2^m) \right] \quad (20f)$$

$$= H(Y^n, Y_3^n | W_3^m, W^m, S_2^m) - H(Y^n, Y_3^n | W_3^m, W^m, S_1^m, S_2^m)$$

$$= I(S_1^m; Y^n, Y_3^n | W_3^m, W^m, S_2^m)$$

$$\geq I(S_1^m; Y^n | W_3^m, W^m, S_2^m) \quad (20g)$$

$$= H(S_1^m | W_3^m, W^m, S_2^m) - H(S_1^m | Y^n, W_3^m, W^m, S_2^m)$$

$$\geq H(S_1^m | W_3^m, W^m, S_2^m) - H(S_1^m | Y^n, W^m, S_2^m) \quad (20h)$$

$$\geq mH(S_1 | S_2, W, W_3) - m\delta(P_e^{(m,n)}), \quad (20i)$$

where (20a) follows from (19); (20b) follows from the fact that $X_{3,1}^k$ is a deterministic function of $(Y_{3,1}^{k-1}, W_3^m)$ and from the memoryless channel assumption, (see (1)); (20c) follows from the fact that conditioning reduces entropy; (20d) follows from causality, [40]; (20e) follows from the fact that $X_{2,k}$ is a deterministic function of S_2^m ; (20f) follows again from the fact that conditioning reduces entropy; (20g) follows from the chain rule for mutual information and the nonnegativity of mutual information; (20h) follows again from the fact that conditioning reduces entropy; and (20i) follows from the memoryless sources and side information assumption and from equations (15)–(16).

Following arguments similar to those that led to (20i) we can also show that

$$\sum_{k=1}^n I(X_{2,k}; Y_k, Y_{3,k} | X_{1,k}, V_k) \geq mH(S_2 | S_1, W, W_3) - m\delta(P_e^{(m,n)}) \quad (21a)$$

$$\sum_{k=1}^n I(X_{1,k}, X_{2,k}; Y_k, Y_{3,k} | V_k) \geq mH(S_1, S_2 | W, W_3) - m\delta(P_e^{(m,n)}). \quad (21b)$$

Next we introduce the time-sharing random variable Q independent of all other random variables. We have $Q = k$ with probability $1/n, k \in \{1, 2, \dots, n\}$. We can write (20i) as

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n I(X_{1,k}; Y_k, Y_{3,k} | X_{2,k}, V_k) &= \frac{1}{n} \sum_{k=1}^n I(X_{1,q}; Y_q, Y_{3,q} | X_{2,q}, V_q, Q = k) \\ &= I(X_{1,Q}; Y_Q, Y_{3,Q} | X_{2,Q}, V_Q) \\ &= I(X_1; Y, Y_3 | X_2, V), \end{aligned} \quad (22)$$

where $X_1 \triangleq X_{1,Q}$, $X_2 \triangleq X_{2,Q}$, $Y \triangleq Y_Q$, $Y_3 \triangleq Y_{3,Q}$ and $V \triangleq (V_Q, Q)$. Since $(X_{1,k}, X_{2,k})$ and $X_{3,k}$ are independent given $V_k = (Y_3^{k-1}, W_3^m)$, for $\bar{v} = (v, k)$ we have

$$\begin{aligned} \Pr\{X_1 = x_1, X_2 = x_2, X_3 = x_3 | V = \bar{v}\} \\ &= \Pr\{X_{1,k} = x_1, X_{2,k} = x_2, X_{3,k} = x_3 | V_k = v, Q = k\} \\ &= \Pr\{X_{1,k} = x_1, X_{2,k} = x_2 | V_k = v, Q = k\} \Pr\{X_{3,k} = x_3 | V_k = v, Q = k\} \\ &= \Pr\{X_1 = x_1, X_2 = x_2 | V = \bar{v}\} \Pr\{X_3 = x_3 | V = \bar{v}\}. \end{aligned} \quad (23)$$

Hence, the probability distribution is of the form given in Thm. 2 for constraints in (13).

From (20)–(22), we have

$$H(S_1 | S_2, W, W_3) - \delta(P_e^{(m,n)}) < \kappa I(X_1; Y, Y_3 | X_2, V) \quad (24a)$$

$$H(S_2 | S_1, W, W_3) - \delta(P_e^{(m,n)}) < \kappa I(X_2; Y, Y_3 | X_1, V) \quad (24b)$$

$$H(S_1, S_2 | W, W_3) - \delta(P_e^{(m,n)}) < \kappa I(X_1, X_2; Y, Y_3 | V). \quad (24c)$$

Finally, taking the limit as $m, n \rightarrow \infty$ and letting $P_e^{(m,n)} \rightarrow 0$ leads to the constraints in (13).

V. OPTIMALITY OF SOURCE-CHANNEL SEPARATION FOR FADING GAUSSIAN MARCS AND MABRCs

In this section we use the separation-based achievable rate and the necessary conditions in (12) to identify conditions for sources transmitted over fading Gaussian MARCs with side information, under which separation is optimal. We begin by considering phase fading Gaussian MARCs defined in (7). The result is stated in the following theorem.

Theorem 4. Consider the transmission of arbitrarily correlated sources S_1 and S_2 over a phase fading Gaussian MARC with receiver side information W and relay side information W_3 . Let the channel inputs be subject to per-symbol power constraints

$$\mathbb{E}\{|X_i|^2\} \leq P_i, \quad i = 1, 2, 3, \quad (25)$$

and let the channel coefficients and the channel input powers satisfy

$$a_{11}^2 P_1 + a_{31}^2 P_3 \leq a_{13}^2 P_1 \quad (26a)$$

$$a_{21}^2 P_2 + a_{31}^2 P_3 \leq a_{23}^2 P_2 \quad (26b)$$

$$a_{11}^2 P_1 + a_{21}^2 P_2 + a_{31}^2 P_3 \leq a_{13}^2 P_1 + a_{23}^2 P_2. \quad (26c)$$

A source-channel rate κ is achievable if

$$H(S_1|S_2, W) < \kappa \log_2(1 + a_{11}^2 P_1 + a_{31}^2 P_3) \quad (27a)$$

$$H(S_2|S_1, W) < \kappa \log_2(1 + a_{21}^2 P_2 + a_{31}^2 P_3) \quad (27b)$$

$$H(S_1, S_2|W) < \kappa \log_2(1 + a_{11}^2 P_1 + a_{21}^2 P_2 + a_{31}^2 P_3). \quad (27c)$$

Conversely, if source-channel rate κ is achievable, then conditions (27) are satisfied with $<$ replaced by \leq .

Proof: See subsections V-A1, V-A2. ■

Remark 6. The source-channel rate κ in Thm. 4 is achieved by using $X_i \sim \mathcal{CN}(0, P_i)$, $i \in \{1, 2, 3\}$, all i.i.d. and independent of each other, and applying DF at the relay.

Remark 7. The achievability proof of the region (8) in Thm. 1 requires strong typicality. Although strong typicality does not apply to continuous random variables, the Markov lemma [41, Lemma 4.1] holds for Gaussian input distributions, [42], therefore we conclude that the achievability of the region (8) is applicable for (27) as well. See also [2, Remark 30].

Remark 8. Note that the phase fading MARC is not degraded in the sense of [11], see also [2, Remark 33].

Remark 9. Remark 6 states that the source-channel rate is achieved by using $X_i \sim \mathcal{CN}(0, P_i)$, $i \in \{1, 2, 3\}$ independent of each other; however, the achievability proof of the region in (8) uses dependent codebooks. For this reason, for the achievability scheme of Thm. 4 we use the channel code construction and decoding rules detailed in [30, Subsection 3.E], which are a special case of Thm. 1 and use independent codebooks. The decoding rules detailed in [30, Subsection 3.E] imply that the destination channel decoder does not use any information provided by the destination source decoder. The only information exchange between the source and channel coders is through the bin indices of the transmitted sequences delivered from the source encoders to the channel encoders and from the channel decoders to the source decoders. Hence, Thm. 4 implies that informational separation is optimal in this scenario.

Next, we consider source transmission over Rayleigh fading MARCs.

Theorem 5. Consider transmission of arbitrarily correlated sources S_1 and S_2 over a Rayleigh fading Gaussian MARC with receiver side information W and relay side information W_3 . Let the channel inputs be subject to

per-symbol power constraints as in (25), and let the channel coefficients and the channel input powers satisfy

$$1 + a_{11}^2 P_1 + a_{31}^2 P_3 \leq \frac{a_{13}^2 P_1}{e^{\frac{1}{a_{13}^2 P_1}} E_1\left(\frac{1}{a_{13}^2 P_1}\right)} \quad (28a)$$

$$1 + a_{21}^2 P_2 + a_{31}^2 P_3 \leq \frac{a_{23}^2 P_2}{e^{\frac{1}{a_{23}^2 P_2}} E_1\left(\frac{1}{a_{23}^2 P_2}\right)} \quad (28b)$$

$$1 + a_{11}^2 P_1 + a_{21}^2 P_2 + a_{31}^2 P_3 \leq \frac{a_{23}^2 P_2 - a_{13}^2 P_1}{\left(e^{\frac{1}{a_{23}^2 P_2}} E_1\left(\frac{1}{a_{23}^2 P_2}\right) - e^{\frac{1}{a_{13}^2 P_1}} E_1\left(\frac{1}{a_{13}^2 P_1}\right)\right)}, \quad (28c)$$

where $E_1(x) \triangleq \int_{q=x}^{\infty} \frac{1}{q} e^{-q} dq$, see [43, Eqn. (5.1.1)]. A source-channel rate κ is achievable if

$$H(S_1|S_2, W) < \kappa \mathbb{E}_{\tilde{U}} \left\{ \log_2(1 + a_{11}^2 |U_{11}|^2 P_1 + a_{31}^2 |U_{31}|^2 P_3) \right\} \quad (29a)$$

$$H(S_2|S_1, W) < \kappa \mathbb{E}_{\tilde{U}} \left\{ \log_2(1 + a_{21}^2 |U_{21}|^2 P_2 + a_{31}^2 |U_{31}|^2 P_3) \right\} \quad (29b)$$

$$H(S_1, S_2|W) < \kappa \mathbb{E}_{\tilde{U}} \left\{ \log_2(1 + a_{11}^2 |U_{11}|^2 P_1 + a_{21}^2 |U_{21}|^2 P_2 + a_{31}^2 |U_{31}|^2 P_3) \right\}. \quad (29c)$$

Conversely, if source-channel rate κ is achievable, then conditions (29) are satisfied with $<$ replaced by \leq .

Proof: The proof uses [30, Corollary 1] and follows similar arguments to those in the proof of Thm. 4. ■

Remark 10. The source-channel rate κ in Thm. 5 is achieved by using $X_i \sim \mathcal{CN}(0, P_i)$, $i \in \{1, 2, 3\}$, all i.i.d. and independent of each other, and applying DF at the relay.

A. Proof of Thm. 4

1) Achievability Proof of Thm. 4:

- *Code construction:* For $i = 1, 2$, assign every $\mathbf{s}_i \in \mathcal{S}_i^m$ to one of 2^{mR_i} bins independently according to a uniform distribution over $\mathcal{U}_i \triangleq \{1, 2, \dots, 2^{mR_i}\}$. Denote these assignments by f_i . A channel code based on DF with rates \hat{R}_1 and \hat{R}_2 , with blocklength n , is constructed as detailed in [30, Subsection 3.E]. Transmitter i can transmit $2^{n\hat{R}_i}$ messages, thus we require $\hat{R}_i = \frac{1}{\kappa} R_i$.
- *Encoding:* Consider a source sequence of length Bm , $\mathbf{s}_i^{Bm} \in \mathcal{S}_i^{Bm}$, $i = 1, 2$. Partition this sequence into B length- m subsequences, $\mathbf{s}_{i,b}$, $b = 1, 2, \dots, B$. Similarly partition the side information sequence $\mathbf{w}^{Bm} \in \mathcal{W}^{Bm}$, into B length- m subsequences. We transmit a total of Bm source samples over $B + 1$ blocks of n channel uses each. If we set $n = \kappa m$, by increasing B we obtain a source-channel rate $(B + 1)n/Bm \rightarrow n/m = \kappa$ as $B \rightarrow \infty$.

Encoding at the transmitters: In block b , $b = 1, 2, \dots, B$, source terminal i , $i = 1, 2$, observes $\mathbf{s}_{i,b}$ and finds its corresponding bin index $u_{i,b} \in \mathcal{U}_i$. Each transmitter sends its corresponding bin index using the channel code described in [30, Subsection 3.E].

Encoding at the relay: Assume that at time b the relay knows $(u_{1,b-1}, u_{2,b-1})$. The relay sends these bin indices using the channel code described in [30, Subsection 3.E].

- *Decoding and error probability:* From [2, Thm. 9] it follows that conditions (26) imply that the achievable channel rate region for decoding at the relay contains the achievable channel rate region for decoding at the

destination. Hence, reliable decoding of the channel code at the destination implies reliable decoding of the channel code at the relay. When the channel coefficients and the channel input powers satisfy the conditions in (26), the RHS of the constraints in (27) characterizes the capacity region of the phase fading Gaussian MARC, (see [2, Thm. 9], [30, Subsection 3.E]). Hence, the transmitted bin indices $\{u_{1,b}, u_{2,b}\}_{b=1}^B$ can be reliably decoded at the destination if,

$$R_1 \leq \kappa \log_2(1 + a_{11}^2 P_1 + a_{31}^2 P_3) \quad (30a)$$

$$R_2 \leq \kappa \log_2(1 + a_{21}^2 P_2 + a_{31}^2 P_3) \quad (30b)$$

$$R_1 + R_2 \leq \kappa \log_2(1 + a_{11}^2 P_1 + a_{21}^2 P_2 + a_{31}^2 P_3). \quad (30c)$$

Decoding the sources at the destination: The decoded bin indices, denoted $(\tilde{u}_{1,b}, \tilde{u}_{2,b}), b = 1, 2, \dots, B$, are given to the source decoder at the destination. Using the bin indices and the side information \mathbf{w}_b , the source decoder at the destination estimates $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$. More precisely, given the bin indices $(\tilde{u}_{1,b}, \tilde{u}_{2,b})$, it declares $(\tilde{\mathbf{s}}_{1,b}, \tilde{\mathbf{s}}_{2,b}) \in S_1^m \times S_2^m$ to be the decoded sequences if it is the unique pair of sequences that satisfies $f_1(\tilde{\mathbf{s}}_{1,b}) = \tilde{u}_{1,b}, f_2(\tilde{\mathbf{s}}_{2,b}) = \tilde{u}_{2,b}$ and $(\tilde{\mathbf{s}}_{1,b}, \tilde{\mathbf{s}}_{2,b}, \mathbf{w}_b) \in A_{\epsilon}^{*(m)}(S_1, S_2, W)$. From the Slepian-Wolf theorem [35, Thm 14.4.1], $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$ can be reliably decoded at the destination if

$$H(S_1|S_2, W) < R_1 \quad (31a)$$

$$H(S_2|S_1, W) < R_2 \quad (31b)$$

$$H(S_1, S_2|W) < R_1 + R_2. \quad (31c)$$

Combining conditions (30) and (31) yields (27) and completes the achievability proof.

2) *Converse Proof of Thm. 4:* Consider the upper bound of Thm. 2. From [2, Thm. 8] it follows that for phase fading with Rx-CSI, the mutual information expressions on the RHS of (12) are simultaneously maximized by X_1, X_2, X_3 independent, zero-mean complex Normal, $X_i \sim \mathcal{CN}(0, P_i), i = 1, 2, 3$, yielding the same expressions as in (27). Therefore, for phase fading MARCs, when conditions (26) hold, the conditions in (27) coincide with those in the upper bound of (12) in Thm. 2, and the conditions in (27) need to be satisfied with \leq instead of $<$.

B. Fading MABRCs

The optimality of informational separation can be established for MABRCs using the results for the MARC with additional constraints, as indicated in the following theorem:

Theorem 6. For phase fading MABRCs for which the conditions in (26) hold together with

$$H(S_1|S_2, W_3) \leq H(S_1|S_2, W) \quad (32a)$$

$$H(S_2|S_1, W_3) \leq H(S_2|S_1, W) \quad (32b)$$

$$H(S_1, S_2|W_3) \leq H(S_1, S_2|W), \quad (32c)$$

the maximum achievable source-channel rate κ satisfies (27). The same statement holds for Rayleigh fading MABRCs with (28) replacing (26) and (29) replacing (27).

Proof: The achievability proof of Thm. 6 differs from the achievability proof of Thm. 4 only due to decoding the source sequences at the relay. Conditions in (26) imply that the reliable decoding of the channel code at the destination lead to the reliable decoding of the channel code at the relay. Conditions (32) imply that the reliable decoding of the source code at the destination lead to the reliable decoding of the source code at the relay, since the relay achievable source rate region contains the destination achievable source rate region. Hence, if conditions (26) and (32) hold, $(s_{1,b}, s_{2,b})$ can be reliably decoded by both the relay and the destination. The converse follows from the upper bound of Thm. 3 and by following similar arguments to the converse proof of Thm. 4. ■

Remark 11. Conditions (32) imply that for the scenario described in Thm. 4 regular encoding and irregular encoding yield the same source-channel achievable rates (see Remark 2); hence, the channel code construction of [30, Subsection 3.E] can be used.

VI. JOINT SOURCE-CHANNEL ACHIEVABLE RATES FOR DISCRETE MEMORYLESS MARCS AND MABRCs

In this section we derive two sets of sufficient conditions for the achievability of source-channel rate $\kappa = 1$ for DM MARCs and MABRCs with correlated sources and side information. Both achievability schemes are established by using a combination of Slepian-Wolf (SW) source coding [38], the CPM technique [9] and DF scheme with successive decoding at the relay and backward decoding at the destination. The techniques differ in the way the source codes are combined. In the first scheme (Thm. 7) the SW code is used for sending information to the destination and CPM is used for sending information to the relay. In the second scheme (Thm. 8), CPM is used for sending information to the destination while SW code is used for sending information to the relay.

Before presenting the results, we first motivate this section by demonstrating sub-optimality of separate encoding for the MARC.

A. Sub-Optimality of Separation for DM MARCs and MABRCs

Consider the transmission of arbitrarily correlated sources S_1 and S_2 over a DM semi-orthogonal MARC (SOMARC) in which the relay-destination link is orthogonal to the channel from the sources to the relay and the destination. The SOMARC setup is characterized by the distribution chain $p(y_R, y_S, y_3|x_1, x_2, x_3) = p(y_R|x_3)p(y_S, y_3|x_1, x_2)$. The SOMARC is depicted in Figure 4. In the following we shall show that joint source-channel code achieves sum rate exceeding the sum-capacity of the SOMARC.

We begin with the outer bound on the sum-capacity of the SOMARC characterized in Proposition 1.

Proposition 1. The sum-capacity of the SOMARC is upper bounded by

$$R_1 + R_2 \leq \max_{p(x_1)p(x_2)p(x_3)} \min \{I(X_1, X_2; Y_3, Y_S), I(X_3; Y_R) + I(X_1, X_2; Y_S)\}. \quad (33)$$

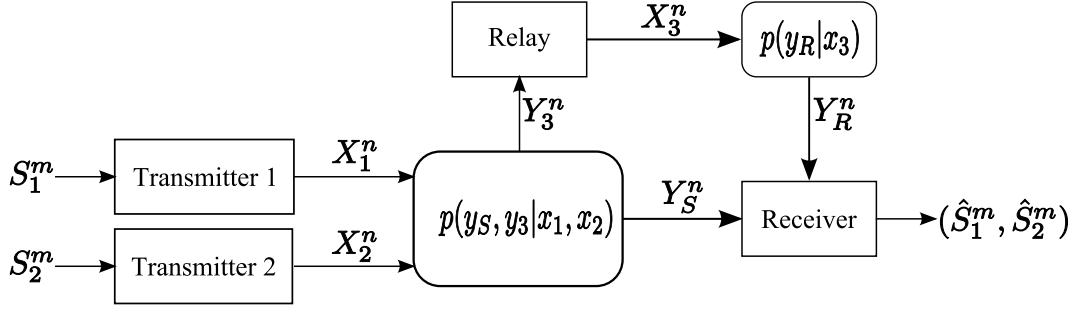


Fig. 4. Semi-orthogonal multiple-access relay channel.

Proof: From the cut-set bound, [35, Thm. 14.10.1] it follows that

$$R_1 + R_2 \leq \max_{p(x_1)p(x_2)p(x_3|x_1,x_2)} \min \{I(X_1, X_2; Y_3, Y_S | X_3), I(X_1, X_2, X_3; Y_R, Y_S)\}. \quad (34)$$

For the mutual information expression $I(X_1, X_2, X_3; Y_R, Y_S)$ we derive the following upper bound

$$I(X_1, X_2, X_3; Y_R, Y_S) = I(X_3; Y_R) + I(X_1, X_2; Y_R | X_3) + I(X_1, X_2; Y_S | Y_R) + I(X_3; Y_S | X_1, X_2, Y_R) \quad (35a)$$

$$= I(X_3; Y_R) + I(X_1, X_2; Y_S | Y_R) + I(X_3; Y_S | X_1, X_2, Y_R) \quad (35b)$$

$$= I(X_3; Y_R) + I(X_1, X_2; Y_S | Y_R) \quad (35c)$$

$$\leq I(X_3; Y_R) + I(X_1, X_2; Y_S), \quad (35d)$$

where (35a) follows from the chain rule for mutual information; (35b) follows from the fact that Y_R and (X_1, X_2) are independent given X_3 ; (35c) follows from the fact that Y_S and X_3 are independent given (X_1, X_2) ; and (35d) follows from the fact that the relay-destination link is orthogonal to the channel from the sources to the destination. From the orthogonality of the relay-destination link we can also conclude that

$$I(X_1, X_2; Y_3, Y_S | X_3) \leq I(X_1, X_2; Y_3, Y_S). \quad (36)$$

Combining (35) and (36) results in (33). Finally note that statistical dependence between (X_1, X_2) and X_3 does not affect the bound expressions, thus, we may set $p(x_3|x_1, x_2) = p(x_3)$ arriving to the distribution chain of Proposition 1. ■

We now consider the transmission of correlated sources (S_1, S_2) with the joint distribution

$$p(s_1 = 0, s_2 = 0) = p(s_1 = 0, s_2 = 1) = p(s_1 = 1, s_2 = 1) = \frac{1}{3}, \quad (37)$$

over a SOMARC defined by

$$\begin{aligned} \mathcal{X}_1 &= \mathcal{X}_1 = \mathcal{X}_3 = \mathcal{Y}_3 = \mathcal{Y}_R = \{0, 1\}, \\ \mathcal{Y}_S &= \{0, 1, 2\}, \\ Y_R &= X_3, \\ Y_3 &= X_1 \oplus X_2, \\ Y_S &= X_1 + X_2. \end{aligned} \quad (38)$$

We use the same sources used in [9] to show the sub-optimality of separation over MAC. For the channel defined in (38) the mutual information expression $I(X_1, X_2; Y_3, Y_S)$ reduces to $I(X_1, X_2; Y_S)$ since $I(X_1, X_2; Y_3, Y_S) = I(X_1, X_2; Y_S) + I(X_1, X_2; Y_3|Y_S)$, but both Y_3 and Y_S are deterministic functions of (X_1, X_2) , and Y_3 is a deterministic function of Y_S . Hence, for independent X_1 and X_2

$$\begin{aligned} R_1 + R_2 &\leq \max_{p(x_1)p(x_2)p(x_3)} \min \{I(X_1, X_2; Y_3, Y_S), I(X_3; Y_R) + I(X_1, X_2; Y_S)\} \\ &= \max_{p(x_1)p(x_2)} \{I(X_1, X_2; Y_S)\} \\ &= 1.5 \text{ bits per channel use.} \end{aligned} \quad (39)$$

Here $H(S_1, S_2) = \log 3 = 1.58$ bits per sample. Hence, we have $H(S_1, S_2) > I(X_1, X_2; Y_S)$, for any joint input distribution of the form $p(x_1)p(x_2)$. Therefore, it is not possible, even by using Slepian-Wolf coding, to send S_1 and S_2 reliably to Y . However, by choosing $X_1 = S_1$ and $X_2 = S_2$ zero error probability is achievable. This example shows that separate source and channel coding is in general sub-optimal for sending arbitrarily correlated sources over MARCs.

B. Joint Source-Channel Coding for MARCs and MABRCs

Thm. 7 and Thm. 8 below present two new sets of sufficient conditions for the achievability of source-channel rate $\kappa = 1$, obtained by combining binning and CPM.

Theorem 7. For DM MARCs and MABRCs with relay and receiver side information as defined in Section II-A, source-channel rate $\kappa = 1$ is achievable if,

$$H(S_1|S_2, W_3) < I(X_1; Y_3|S_2, X_2, V_1, X_3, W_3) \quad (40a)$$

$$H(S_2|S_1, W_3) < I(X_2; Y_3|S_1, X_1, V_2, X_3, W_3) \quad (40b)$$

$$H(S_1, S_2|W_3) < I(X_1, X_2; Y_3|V_1, V_2, X_3, W_3) \quad (40c)$$

$$H(S_1|S_2, W) < I(X_1, X_3; Y|S_1, X_2, V_2) \quad (40d)$$

$$H(S_2|S_1, W) < I(X_2, X_3; Y|S_2, X_1, V_1) \quad (40e)$$

$$H(S_1, S_2|W) < I(X_1, X_2, X_3; Y|S_1, S_2), \quad (40f)$$

for a joint distribution that factors as

$$p(s_1, s_2, w_3, w)p(v_1)p(x_1|s_1, v_1)p(v_2)p(x_2|s_2, v_2)p(x_3|v_1, v_2)p(y_3, y|x_1, x_2, x_3). \quad (41)$$

Proof: The proof is given in Subsection VI-D. ■

Theorem 8. For DM MARCs and MABRCs with relay and receiver side information as defined in Section II-A,

source-channel rate $\kappa = 1$ is achievable if,

$$H(S_1|S_2, W_3) < I(X_1; Y_3|S_1, X_2, X_3) \quad (42a)$$

$$H(S_2|S_1, W_3) < I(X_2; Y_3|S_2, X_1, X_3) \quad (42b)$$

$$H(S_1, S_2|W_3) < I(X_1, X_2; Y_3|S_1, S_2, X_3) \quad (42c)$$

$$H(S_1|S_2, W) < I(X_1, X_3; Y|S_2, X_2, W) \quad (42d)$$

$$H(S_2|S_1, W) < I(X_2, X_3; Y|S_1, X_1, W) \quad (42e)$$

$$H(S_1, S_2|W) < I(X_1, X_2, X_3; Y|W), \quad (42f)$$

for a joint distribution that factors as

$$p(s_1, s_2, w_3, w)p(x_1|s_1)p(x_2|s_2)p(x_3|s_1, s_2)p(y_3, y|x_1, x_2, x_3). \quad (43)$$

Proof: The proof is given in Subsection VI-E. ■

C. Discussion

In Figure 5 we illustrate the Markov chains for the joint distribution considered in Thm. 7 and Thm. 8.

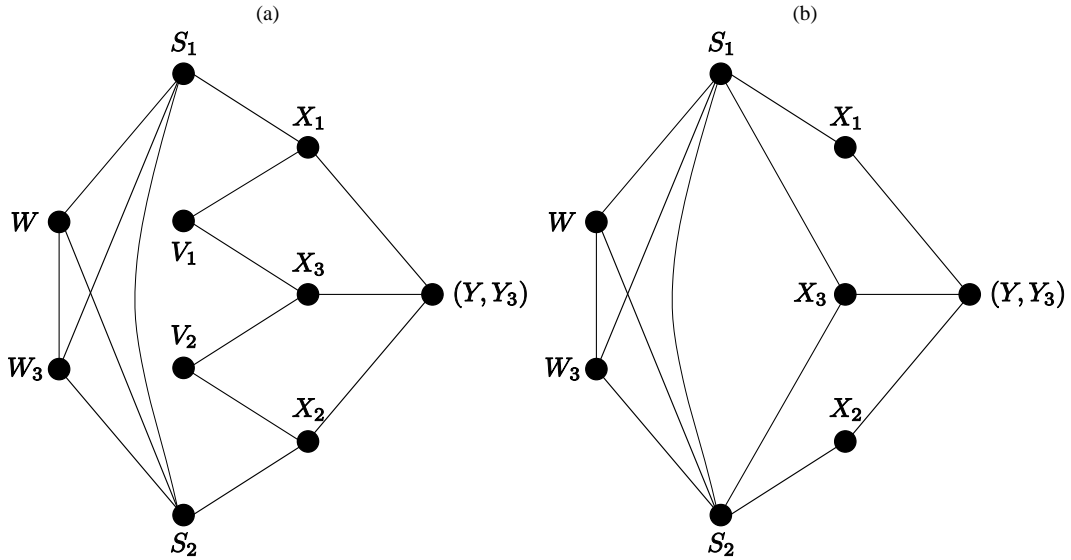


Fig. 5. (a) Diagram of the Markov chain for the joint distribution considered in (41); (b) Diagram of the Markov chain for the joint distribution considered in (43).

Remark 12. In Thm. 7 and Thm. 8, conditions (40a)–(40c) and (42a)–(42c) respectively, are constraints for decoding at the relay, while conditions (40d)–(40f) and (42d)–(42f), are decoding constraints at the destination.

Remark 13. Consider Thm. 7: the LHS of condition (40a) is the residual entropy of S_1 when (S_2, W_3) are known. In the RHS of (40a), as V_1, S_2, X_2, X_3 and W_3 are given, the mutual information expression $I(X_1; Y_3|S_2, X_2, V_1, X_3, W_3)$

represents the available rate that can be used to send information on the *source sequence* S_1^m . The LHS of condition (40d) is the entropy of S_1 when (S_2, W) are known. The RHS of condition (40d) expresses the amount of *binning information* that can be reliably transmitted from sender 1 and the relay to the destination. This follows as the mutual information expression in (40d) can be written as $I(X_1, X_3; Y|S_1, X_2, V_2) = I(X_1, X_3; Y|S_1, S_2, V_2, X_2, W)$, which represents the rate that can be used to send the *bin index* u_1 , of source sequence S_1^m , to the destination (see subsection VI-D). This is in contrast to the decoding constraint of the relay, cf. (40a), which represents the *source information* S_1 transmitted *in excess of the bin index* sent to the destination. Therefore each mutual information expression in (40a) and (40d), represents *different* types of information sent by the source: either source-channel codeword to the relay (40a) or bin index to the destination (40d). This is because SW is used for the destination and CPM is used for the relay.

Consider Thm. 8: the LHS of condition (42a) is the residual entropy of S_1 when (S_2, W_3) are known. In the RHS of (42a) the mutual information expression $I(X_1; Y_3|S_1, X_2, X_3) = I(X_1; Y_3|S_1, S_2, X_2, X_3, W_3)$ represents the rate that can be used to send the *bin index* u_1 , of source sequence S_1^m to the relay (see subsection VI-E). The LHS of condition (42d) is the residual entropy of S_1 when (S_2, W) are known. In the RHS of condition (42d), as S_2, X_2 and W are given, the mutual information expression $I(X_1, X_3; Y|S_2, X_2, W)$ represents the available rate that can be used to send information on the *source sequence* S_1^m .

Remark 14. For an input distribution

$$p(s_1, s_2, w_3, w, v_1, v_2, x_1, x_2, x_3) = p(s_1, s_2, w_3, w)p(v_1)p(x_1|v_1)p(v_2)p(x_2|v_2)p(x_3|v_1, v_2),$$

conditions (40) reduce to conditions (8), and the transmission scheme reduces to a separation-based achievable scheme for $\kappa = 1$.

Remark 15. In both Thm. 7 and Thm. 8 the conditions stemming from the CPM technique can be reduced to the MAC source channel conditions of [9, Equation (12)]. In Thm. 7 letting $\mathcal{V}_1 = \mathcal{V}_2 = \mathcal{X}_3 = \mathcal{W}_3 = \phi$, reduces the relay conditions in (40a)–(40c) to the ones in [9] with Y_3 as the destination. In Thm. 8 letting $\mathcal{X}_3 = \mathcal{W} = \phi$, reduces the destination conditions in (40d)–(40f) to the ones in [9] with Y as the destination.

Remark 16. Thm. 7 and Thm. 8 establish *different* achievability conditions. As stated in Remark 15 and in Subsection VI-A, separate source and channel coding is generally sub-optimal for sending correlated sources over DM MARCs and MABRCs. In Thm. 7 the CPM technique is used for sending information to the relay, while in Thm. 8 a SW code is used for sending information to the relay. This observation implies that the relay decoding constraints of Thm. 7 are looser compared to the relay decoding constraints of Thm. 8. Using similar arguments we conclude that the destination decoding constraints of Thm. 8 are looser compared to the destination decoding constraints of Thm. 7. The above observations indicate a trade-off between the two achievability schemes.

Figure 6 depicts the cooperative relay broadcast channel (CRBC) model, which is a special case of a MARC with a single source terminal. For the CRBC with correlated relay and destination side information, we can identify exactly the optimal source-channel rate, which was previously obtained in [22].

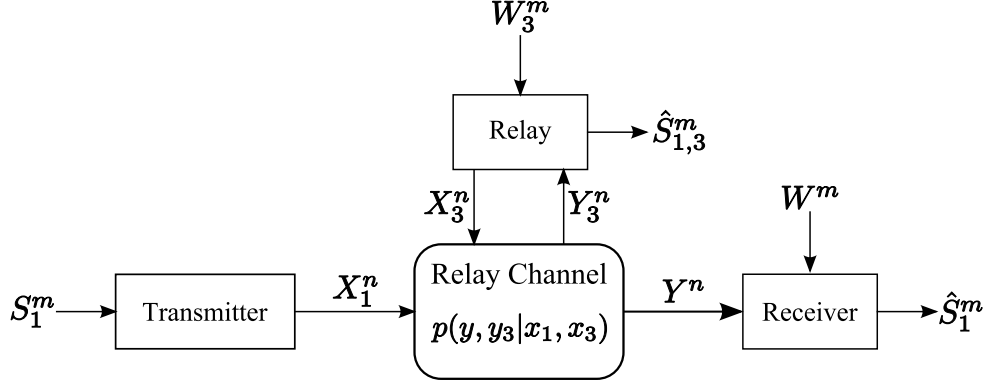


Fig. 6. Cooperative relay broadcast channel with correlated side information.

Corollary 1. For CRBC with relay and receiver side information, source-channel rate κ is achievable if

$$H(S_1|W_3) < \kappa I(X_1; Y_3|X_3) \quad (44a)$$

$$H(S_1|W) < \kappa I(X_1, X_3; Y), \quad (44b)$$

for some input distribution $p(s_1, w_3, w)p(x_1, x_3)$. Conversely, if rate κ is achievable then the conditions in (44a) and (44b) are satisfied with $<$ replaced by \leq for some input distribution $p(x_1, x_3)$.

Proof: The achievability follows from Thm. 1 by assigning $X_3 = V_1$ and $\mathcal{S}_2 = \mathcal{X}_2 = \mathcal{V}_2 = \phi$. The converse follows from Thm. 3. ■

Remark 17. For source-channel rate $\kappa = 1$, the conditions in (44) can also be obtained from Thm. 7 by letting $V_1 = X_3$, $\mathcal{S}_2 = \mathcal{X}_2 = \mathcal{V}_2 = \phi$ and considering an input distribution independent of the sources. However, if we consider the conditions in Thm. 8 for $\mathcal{S}_2 = \mathcal{X}_2 = \phi$, we obtain the following conditions for the achievability

$$H(S_1|W_3) < I(X_1; Y_3|X_3, S_1) \quad (45a)$$

$$H(S_1|W) < I(X_1, X_3; Y|W), \quad (45b)$$

for some input distribution $p(s_1, w_3, w)p(x_1|s_1)p(x_3|s_1)$.

Note that the RHS of inequalities in (45a) and (45b) are smaller than those in (44a) and (44b), respectively. Moreover, not all joint input distributions can be achieved. Hence, the conditions obtained from Thm. 8 for the CRBC setup are stricter than those obtained from Thm. 1, illustrating the fact that the two sets of conditions are not equivalent.

D. Proof of Thm. 7

1) *Code construction:* For $i = 1, 2$, assign every $\mathbf{s}_i \in \mathcal{S}_i^n$ to one of 2^{nR_i} bins independently according to a uniform distribution on $\mathcal{U}_i \triangleq \{1, 2, \dots, 2^{nR_i}\}$. Denote this assignment by $f_i, i = 1, 2$.

For the channel codebook, for each $i = 1, 2$, generate 2^{nR_i} codewords $\mathbf{v}_i(u_i), u_i \in \mathcal{U}_i$, by choosing the letters $v_{i,k}(u_i), k = 1, 2, \dots, n$, independently with distribution $p_{V_i}(v_{i,k})$. For each pair $(\mathbf{s}_i, u_i), i = 1, 2$ generate

| Node | Block 1 | Block 2 | Block B | Block $B + 1$ |
|--------|-------------------------------------|---|---|---------------------------------------|
| User 1 | $\mathbf{s}_{1,1}$ | $\mathbf{s}_{1,2}$ | $\mathbf{s}_{1,B}$ | — |
| | — | $f_1(\mathbf{s}_{1,1}) = u_{1,1}$ | $f_1(\mathbf{s}_{1,B-1}) = u_{1,B-1}$ | $f_1(\mathbf{s}_{1,B}) = u_{1,B}$ |
| | $\mathbf{x}_1(\mathbf{s}_{1,1}, 1)$ | $\mathbf{x}_1(\mathbf{s}_{1,2}, u_{1,1})$ | $\mathbf{x}_1(\mathbf{s}_{1,B}, u_{1,B-1})$ | $\mathbf{x}_1(\mathbf{a}_1, u_{1,B})$ |
| | $\mathbf{v}_1(1)$ | $\mathbf{v}_1(u_{1,1})$ | $\mathbf{v}_1(u_{1,B-1})$ | $\mathbf{v}_1(u_{1,B})$ |
| User 2 | $\mathbf{s}_{2,1}$ | $\mathbf{s}_{2,2}$ | $\mathbf{s}_{2,B}$ | — |
| | — | $f_2(\mathbf{s}_{2,1}) = u_{2,1}$ | $f_2(\mathbf{s}_{2,B-1}) = u_{2,B-1}$ | $f_2(\mathbf{s}_{2,B}) = u_{2,B}$ |
| | $\mathbf{x}_2(\mathbf{s}_{2,1}, 1)$ | $\mathbf{x}_2(\mathbf{s}_{2,2}, u_{2,1})$ | $\mathbf{x}_2(\mathbf{s}_{2,B}, u_{2,B-1})$ | $\mathbf{x}_2(\mathbf{a}_2, u_{2,B})$ |
| | $\mathbf{v}_2(1)$ | $\mathbf{v}_2(u_{2,1})$ | $\mathbf{v}_2(u_{2,B-1})$ | $\mathbf{v}_2(u_{2,B})$ |
| Relay | $\mathbf{x}_3(1, 1)$ | $\mathbf{x}_3(u_{1,1}, u_{2,1})$ | $\mathbf{x}_3(u_{1,B-1}, u_{2,B-1})$ | $\mathbf{x}_3(u_{1,B}, u_{2,B})$ |

TABLE II
ENCODING PROCESS FOR THE JOINT SOURCE-CHANNEL ACHIEVABLE RATE OF THM. 7.

one n -length codeword $\mathbf{x}_i(\mathbf{s}_i, u_i)$, $\mathbf{s}_i \in \mathcal{S}_i^n$, by choosing the letters $x_{i,k}(\mathbf{s}_i, u_i)$ independently with distribution $p_{X_i|S_i, V_i}(x_{i,k}|s_{i,k}, v_{i,k}(u_i))$ for all $1 \leq k \leq n$. Finally, generate one length- n relay codeword $\mathbf{x}_3(u_1, u_2)$ for each pair $(u_1, u_2) \in \mathcal{U}_1 \times \mathcal{U}_2$ by choosing $x_{3,k}(u_1, u_2)$ independently with distribution $p_{X_3|V_1, V_2}(x_{3,k}|v_{1,k}(u_1), v_{2,k}(u_2))$ for all $1 \leq k \leq n$.

2) *Encoding*: (See Table II). Consider a source sequence $\mathbf{s}_i^{Bn} \in \mathcal{S}_i^{Bn}$, $i = 1, 2$ of length Bn . Partition this sequence into B length- n subsequences, $\mathbf{s}_{i,b}$, $b = 1, \dots, B$. Similarly, for $b = 1, 2, \dots, B$, partition the side information sequences w_3^{Bn} and w^{Bn} into B length- n subsequences $\mathbf{w}_{3,b}, \mathbf{w}_b$ respectively. We transmit a total of Bn source samples over $B + 1$ blocks of n channel uses each. In block 1, source terminal i , $i = 1, 2$, transmits the channel codeword $\mathbf{x}_i(\mathbf{s}_{i,1}, 1)$. In block b , $b = 2, \dots, B$, source terminal i , $i = 1, 2$, transmits the channel codeword $\mathbf{x}_i(\mathbf{s}_{i,b}, u_{i,b-1})$ where $u_{i,b-1} \in \mathcal{U}_i$ is the bin index of source vector $\mathbf{s}_{i,b-1}$. In block $B + 1$, source terminal i , $i = 1, 2$, transmits $\mathbf{x}_i(\mathbf{a}_i, u_{i,B})$ where $\mathbf{a}_i \in \mathcal{S}_i^n$ is a fixed sequence.

At block $b = 1$, the relay simply transmits $\mathbf{x}_3(1, 1)$. Assume that at block b , $b = 2, \dots, B, B + 1$, the relay estimates $(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$. It then finds the corresponding bin indices $\hat{u}_{i,b-1} \in \mathcal{U}_i$, $i = 1, 2$, and transmits the channel codeword $\mathbf{x}_3(\hat{u}_{1,b-1}, \hat{u}_{2,b-1})$.

3) *Decoding*: The relay decodes the source sequences sequentially trying to reconstruct source block $\mathbf{s}_{i,b}$, $i = 1, 2$, at the end of channel block b as follows: assume that the relay knows $(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$ at the end of block $b - 1$. Hence, it can find the bin indices $(u_{1,b-1}, u_{2,b-1})$. Using this information, its received signal $\mathbf{y}_{3,b}$ and the side information $\mathbf{w}_{3,b}$, the relay decodes $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$, by looking for a unique pair $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ such that:

$$\begin{aligned}
&(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_1(\tilde{\mathbf{s}}_1, u_{1,b-1}), \mathbf{x}_2(\tilde{\mathbf{s}}_2, u_{2,b-1}), \mathbf{v}_1(u_{1,b-1}), \mathbf{v}_2(u_{2,b-1}), \mathbf{x}_3(u_{1,b-1}, u_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \\
&\in A_{\epsilon}^{*(n)}(S_1, S_2, X_1, X_2, V_1, V_2, X_3, W_3, Y_3).
\end{aligned} \tag{46}$$

Decoding at the destination is done using backward decoding. The destination node waits until the end of channel block $B + 1$. It first tries to decode $(\mathbf{s}_{1,B}, \mathbf{s}_{2,B})$ using the received signal at channel block $B + 1$ and its

side information \mathbf{w}_B . Going backwards from the last channel block to the first, we assume that the destination knows $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ and consider decoding of $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$. At block $b + 1$ the destination channel decoder first estimates the destination bin indices $u_{i,b}, i = 1, 2$, corresponding to $\mathbf{s}_{i,b}$ based on its received signal \mathbf{y}_{b+1} and the side information \mathbf{w}_{b+1} . More precisely, the destination channel decoder looks for a unique pair $(\tilde{u}_1, \tilde{u}_2)$ such that:

$$\begin{aligned} & (\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1}, \mathbf{x}_1(\mathbf{s}_{1,b+1}, \tilde{u}_1), \mathbf{x}_2(\mathbf{s}_{2,b+1}, \tilde{u}_2), \mathbf{v}_1(\tilde{u}_1), \mathbf{v}_2(\tilde{u}_2), \mathbf{x}_3(\tilde{u}_1, \tilde{u}_2), \mathbf{w}_{b+1}, \mathbf{y}_{b+1}) \\ & \in A_\epsilon^{*(n)}(S_1, S_2, X_1, X_2, V_1, V_2, X_3, W, Y). \end{aligned} \quad (47)$$

The decoded destination bin indices, $(\tilde{u}_1, \tilde{u}_2)$ are then given to the destination source decoder. From the destination bin indices and the side information \mathbf{w}_b , the destination source decoder estimates $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$. More precisely, given the destination bin indices $(\tilde{u}_1, \tilde{u}_2)$, the destination source decoder declares $(\tilde{\mathbf{s}}_{1,b}, \tilde{\mathbf{s}}_{2,b})$ as the decoded sequences if it is the unique pair of sequences that satisfies $f_1(\tilde{\mathbf{s}}_{1,b}) = \tilde{u}_1, f_2(\tilde{\mathbf{s}}_{2,b}) = \tilde{u}_2$ and $(\tilde{\mathbf{s}}_{1,b}, \tilde{\mathbf{s}}_{2,b}, \mathbf{w}_b) \in A_\epsilon^{*(n)}(S_1, S_2, W)$.

4) *Error probability analysis and corresponding rate constraints:* The error probability analysis is given in Appendix C. It is shown that relay decoding (46) can be done reliably if (40a)–(40c) hold and decoding the source sequences at the destination can be done reliably as long as (40d)–(40f) hold.

E. Proof of Thm. 8

1) *Code construction:* For $i = 1, 2$, assign every $\mathbf{s}_i \in \mathcal{S}_i^n$ to one of 2^{nR_i} bins independently according to a uniform distribution on $\mathcal{U}_i \triangleq \{1, 2, \dots, 2^{nR_i}\}$. Denote this assignment by $f_i, i = 1, 2$.

For $i = 1, 2$, for each pair $(u_i, \mathbf{s}_i), u_i \in \mathcal{U}_i, \mathbf{s}_i \in \mathcal{S}_i^m$, generate one n -length codeword $\mathbf{x}_i(u_i, \mathbf{s}_i)$, by choosing the letters $x_{i,k}(u_i, \mathbf{s}_i)$ independently with distribution $p_{X_i|S_i}(x_{i,k}|\mathbf{s}_{i,k})$ for all $1 \leq k \leq n$. Finally, generate one length- n relay codeword $\mathbf{x}_3(\mathbf{s}_1, \mathbf{s}_2)$ for each pair $(\mathbf{s}_1, \mathbf{s}_2) \in \mathcal{S}_1^n \times \mathcal{S}_2^n$ by choosing $x_{3,k}(\mathbf{s}_1, \mathbf{s}_2)$ independently with distribution $p_{X_3|S_1, S_2}(x_{3,k}|\mathbf{s}_{1,k}, \mathbf{s}_{2,k})$ for all $1 \leq k \leq n$.

2) *Encoding:* (See Table III). Consider a source sequence $\mathbf{s}_{i,1}^{Bn} \in \mathcal{S}_i^{Bn}, i = 1, 2$ of length Bn . Partition this sequence into B length- n subsequences, $\mathbf{s}_{i,b}, b = 1, \dots, B$. Similarly, for $b = 1, 2, \dots, B$, partition the side information sequences \mathbf{w}_3^{Bn} and \mathbf{w}_b^{Bn} into B length- n subsequences $\mathbf{w}_{3,b}, \mathbf{w}_b$ respectively. We transmit a total of Bn source samples over $B + 1$ blocks of n channel uses each. In block 1, source terminal $i, i = 1, 2$, observes $\mathbf{s}_{i,1}$ and finds its corresponding bin index $u_{i,1} \in \mathcal{U}_i$. It transmits the channel codeword $\mathbf{x}_i(u_{i,1}, \mathbf{a}_i)$ where $\mathbf{a}_i \in \mathcal{S}_i^n$ is a fixed sequence. In block $b, b = 2, \dots, B$, source terminal $i, i = 1, 2$, transmits the channel codeword $\mathbf{x}_i(u_{i,b}, \mathbf{s}_{i,b-1})$ where $u_{i,b} \in \mathcal{U}_i$ is the bin index of source vector $\mathbf{s}_{i,b}$. In block $B + 1$, source terminal $i, i = 1, 2$, transmits $\mathbf{x}_i(1, \mathbf{s}_{i,B})$.

At block $b = 1$, the relay simply transmits $\mathbf{x}_3(\mathbf{a}_1, \mathbf{a}_2)$. Assume that at block $b, b = 2, \dots, B, B + 1$, the relay estimates $\hat{\mathbf{s}}_{i,b-1}, i = 1, 2$. It then transmits the channel codeword $\mathbf{x}_3(\hat{\mathbf{s}}_{1,b-1}, \hat{\mathbf{s}}_{2,b-1})$.

3) *Decoding:* The relay decodes the source sequences sequentially trying to reconstruct source block $\mathbf{s}_{i,b}, i = 1, 2$, at the end of channel block b as follows: assume that the relay knows $\mathbf{s}_{i,b-1}, i = 1, 2$, at the end of block $b - 1$. Using this information, its received signal $\mathbf{y}_{3,b}$, and the side information $\mathbf{w}_{3,b-1}$, the relay channel decoder at time

| Node | Block 1 | Block 2 | Block B | Block $B + 1$ |
|--------|--|--|--|--|
| User 1 | $\mathbf{s}_{1,1}$ | $\mathbf{s}_{1,2}$ | $\mathbf{s}_{1,B}$ | — |
| | $f_1(\mathbf{s}_{1,1}) = u_{1,1}$ | $f_1(\mathbf{s}_{1,2}) = u_{1,2}$ | $f_1(\mathbf{s}_{1,B}) = u_{1,B}$ | — |
| | $\mathbf{x}_1(u_{1,1}, \mathbf{a}_1)$ | $\mathbf{x}_1(u_{1,2}, \mathbf{s}_{1,1})$ | $\mathbf{x}_1(u_{1,B}, \mathbf{s}_{1,B-1})$ | $\mathbf{x}_1(1, \mathbf{s}_{1,B})$ |
| User 2 | $\mathbf{s}_{2,1}$ | $\mathbf{s}_{2,2}$ | $\mathbf{s}_{2,B}$ | — |
| | $f_1(\mathbf{s}_{2,1}) = u_{2,1}$ | $f_1(\mathbf{s}_{2,2}) = u_{2,2}$ | $f_1(\mathbf{s}_{2,B}) = u_{2,B}$ | — |
| | $\mathbf{x}_2(u_{2,1}, \mathbf{a}_2)$ | $\mathbf{x}_2(u_{2,2}, \mathbf{s}_{2,1})$ | $\mathbf{x}_2(u_{2,B}, \mathbf{s}_{2,B-1})$ | $\mathbf{x}_2(1, \mathbf{s}_{2,B})$ |
| Relay | $\mathbf{x}_3(\mathbf{a}_1, \mathbf{a}_2)$ | $\mathbf{x}_3(\mathbf{s}_{1,1}, \mathbf{s}_{2,1})$ | $\mathbf{x}_3(\mathbf{s}_{1,B-1}, \mathbf{s}_{2,B-1})$ | $\mathbf{x}_3(\mathbf{s}_{1,B}, \mathbf{s}_{2,B})$ |

TABLE III
ENCODING PROCESS FOR THE JOINT SOURCE-CHANNEL ACHIEVABLE RATE OF THM. 8.

b decodes $(u_{1,b}, u_{2,b})$, i.e., the bin indices corresponding to $\mathbf{s}_{i,b}$, $i = 1, 2$, by looking for a unique pair $(\tilde{u}_1, \tilde{u}_2)$ such that:

$$(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}, \mathbf{x}_1(\tilde{u}_1, \mathbf{s}_{1,b-1}), \mathbf{x}_2(\tilde{u}_2, \mathbf{s}_{2,b-1}), \mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}), \mathbf{w}_{3,b-1}, \mathbf{y}_{3,b}) \in A_{\epsilon}^{*(n)}(S_1, S_2, X_1, X_2, X_3, W_3, Y_3). \quad (48)$$

The decoded bin indices, $(\tilde{u}_1, \tilde{u}_2)$, are then given to the relay source decoder. Using the bin indices and the side information $\mathbf{w}_{3,b}$, the relay source decoder estimates $\mathbf{s}_{1,b}, \mathbf{s}_{2,b}$. More precisely, given the bin indices $(\tilde{u}_1, \tilde{u}_2)$ the relay source decoder declares $(\tilde{\mathbf{s}}_{1,b}, \tilde{\mathbf{s}}_{2,b})$ as the decoded sequences if it is the unique pair of sequences that satisfies $f_1(\tilde{\mathbf{s}}_{1,b}) = \tilde{u}_1, f_2(\tilde{\mathbf{s}}_{2,b}) = \tilde{u}_2$ and $(\tilde{\mathbf{s}}_{1,b}, \tilde{\mathbf{s}}_{2,b}, \mathbf{w}_{3,b}) \in A_{\epsilon}^{*(n)}(S_1, S_2, W_3)$.

Decoding at the destination is done using backward decoding. The destination node waits until the end of channel block $B + 1$. It first tries to decode $\mathbf{s}_{i,B}$, $i = 1, 2$, using the received signal at channel block $B + 1$ and its side information \mathbf{w}_B . Going backwards from the last channel block to the first, we assume that the destination knows $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ and consider decoding of $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$. From $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ the destination can find the bin indices $(u_{1,b+1}, u_{2,b+1})$. Using this information, its received signal \mathbf{y}_{b+1} and the side information \mathbf{w}_b , the destination decodes $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$ by looking for a unique pair $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ such that:

$$(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_1(u_{1,b+1}, \tilde{\mathbf{s}}_1), \mathbf{x}_2(u_{2,b+1}, \tilde{\mathbf{s}}_2), \mathbf{x}_3(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2), \mathbf{w}_b, \mathbf{y}_{b+1}) \in A_{\epsilon}^{*(n)}(S_1, S_2, X_1, X_2, X_3, W, Y). \quad (49)$$

4) *Error probability analysis and corresponding rate constraints:* The error probability analysis is given in Appendix D. It is shown that decoding the source sequences at the relay can be done reliably as long as (42a)–(42c) hold and destination decoding (49) can be done reliably if (42d)–(42f) hold.

Remark 18. In both Thm. 7 and Thm. 8 we used a combination of SW coding and CPM. Since CPM can generally allow higher rates, a natural question that arises is whether it is possible to design a scheme based only on CPM, namely encode both cooperation (relay) information and the new information using a superposition CPM scheme. The problem with this approach in the framework of the current paper, which uses decoding based on joint typicality, is that joint typicality does not apply to different blocks of the same RV. For example, we cannot test the joint

typicality of \mathbf{s}_b and \mathbf{s}_{b+1} , as they belong to different time blocks. Using a CPM only scheme will require us to carry such test. For example, using the CPM technique for sending information to both the relay and the destination yields the following relay decoding rule: assume that the relay knows $(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$ at the end of block $b-1$. The relay decodes $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$, by looking for a unique pair $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ such that:

$$(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_1(\tilde{\mathbf{s}}_1, \mathbf{s}_{1,b-1}), \mathbf{x}_2(\tilde{\mathbf{s}}_2, \mathbf{s}_{2,b-1}), \mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}, \mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)}. \quad (50)$$

Note that $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ and $(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$ can not be jointly typical since they correspond to different blocks indices, i.e. $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ correspond to block b while $(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$ correspond to block $b-1$.

Similarly, the following destination decoding rule becomes: assume that the destination knows $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ at the end of block $b+1$. The destination decodes $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b})$, by looking for a unique pair $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ such that:

$$(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1}, \mathbf{x}_1(\mathbf{s}_{1,b+1}, \tilde{\mathbf{s}}_1), \mathbf{x}_2(\mathbf{s}_{2,b+1}, \tilde{\mathbf{s}}_2), \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_3(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2), \mathbf{w}_b, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)}. \quad (51)$$

Again, $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ and $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ can not be jointly typical since they correspond to different block indices, i.e. $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ correspond to block b while $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ correspond to block $b+1$.

We conclude that applying the CPM technique for sending information to both the relay and the destination can not be used together with a joint typicality decoder. It is, of course, possible to construct schemes that use a different decoder or carry CPM through intermediate RVs that overcome this difficulty, but these are left for future research.

Remark 19. A comparison of the decoding rules of Thm. 7 with the decoding rules of Thm. 8 reveals a difference in side information block indices at the relay and at the destination. The decoding rules of Thm. 7 uses side information block with the same index as that of the received vector, while the decoding rules of Thm. 8 uses side information block index earlier than that of the received vector. The difference stems from the fact that in Thm. 7 the random binning (auxiliary variables) is used for cooperation between the relay and the destination, while in Thm. 8 the cooperation is based on the source sequences. In the DF scheme cooperation information is used with delay. As it is used with delay when source sequences are used for cooperation the side information block should be correlated with the sources blocks.

Remark 20. Both achievability schemes of Thm. 7 and Thm. 8 use joint typicality decoders. The joint typicality decoders can be used only for $\kappa = 1$ (i.e. $n = m$). For $\kappa \neq 1$ \mathbf{x}_1 and \mathbf{s}_1 can not be jointly typical. Hence, a separation based joint typicality decoders should be used.

VII. CONCLUSIONS

In this paper we have considered the transmission of arbitrarily correlated sources over MARCs and MABRCs with correlated side information at both the relay and the destination. We have first derived an operational separation based achievable source-channel rate for MARCs (this rate applies directly to MABRCs as well). This result is established by using an irregular encoding scheme for the channel code. We have also showed that for both MABRCs and MARCs regular encoding is more restrictive than irregular encoding.

Next, we have considered phase and Rayleigh fading MARCs with side information and we have identified the conditions under which informational separation is optimal in this setup. Conditions for the optimality of informational separation for fading MABRCs have also been obtained. The importance of this result lies in the fact that it enables a modular system design (separate design of the source and channel codes) while achieving the optimal end-to-end performance. We also note here that this is the first time the optimality of separation is shown for the MARC and MABRC configurations.

Finally, we have considered joint source-channel coding for DM MARCs and MABRCs for source-channel rate $\kappa = 1$. We first showed with an explicit example that joint source-channel codes in general achieve higher rates for this scenario compared to separation based codes. We then derived two new joint source-channel achievability schemes for DM MARCs and MABRCs. Both schemes use a combination of binning and CPM. While in the first scheme joint source-channel coding is used for encoding information to the relay and binning is used for encoding information to the destination, in the second scheme binning is used for encoding information to the relay and source-channel coding is used for encoding information to the destination. The different combinations of binning and source mapping enable flexibility in the system design by choosing one of the two schemes according to the quality of the side information and received signals at the relay and the destination. In particular, the first scheme has looser decoding constraints at the relay and is therefore better when the source-relay communication is the bottleneck, while the second scheme has looser decoding constraints at the destination, and is more suitable to scenarios where the source-destination path is more noisy.

APPENDIX A

ERROR PROBABILITY ANALYSIS FOR THM. 1

Let ϵ_1 be a positive number such that $\epsilon_1 > \epsilon$, and $\epsilon_1 \rightarrow 0$ as $\epsilon \rightarrow 0$. We start with the relay error probability analysis.

Relay channel decoder: Assuming correct decoding at block $b - 1$, a relay channel decoding error event, E_{ch}^r , in block b , is the union of the following events:

$$\begin{aligned} E_1^r &\triangleq \{(\mathbf{x}_1(u_{1,b}^r, u_{1,b-1}^d), \mathbf{x}_2(u_{2,b}^r, u_{2,b-1}^d), \mathbf{v}_1(u_{1,b-1}^d), \mathbf{v}_2(u_{2,b-1}^d), \mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d), \mathbf{y}_{3,b}) \notin A_\epsilon^{*(n)}\}, \\ E_2^r &\triangleq \{\exists \tilde{u}_1 \neq u_{1,b}^r : (\mathbf{x}_1(\tilde{u}_1, u_{1,b-1}^d), \mathbf{x}_2(u_{2,b}^r, u_{2,b-1}^d), \mathbf{v}_1(u_{1,b-1}^d), \mathbf{v}_2(u_{2,b-1}^d), \mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d), \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)}\}, \\ E_3^r &\triangleq \{\exists \tilde{u}_2 \neq u_{2,b}^r : (\mathbf{x}_1(u_{1,b}^r, u_{1,b-1}^d), \mathbf{x}_2(\tilde{u}_2, u_{2,b-1}^d), \mathbf{v}_1(u_{1,b-1}^d), \mathbf{v}_2(u_{2,b-1}^d), \mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d), \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)}\}, \\ E_4^r &\triangleq \{\exists \tilde{u}_1 \neq u_{1,b}^r, \tilde{u}_2 \neq u_{2,b}^r : \\ &\quad (\mathbf{x}_1(\tilde{u}_1, u_{1,b-1}^d), \mathbf{x}_2(\tilde{u}_2, u_{2,b-1}^d), \mathbf{v}_1(u_{1,b-1}^d), \mathbf{v}_2(u_{2,b-1}^d), \mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d), \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)}\}. \end{aligned}$$

Given correct decoding at block $b - 1$, the average probability of channel decoding error at the relay in block b ,

$\bar{P}_{r,ch}^{(n)2}$, is defined by

$$\bar{P}_{r,ch}^{(n)} = \Pr \left\{ E_1^r \bigcup \bigcup_{j=2}^4 \{ E_j^r \cap (E_1^r)^c \} \right\} \leq \Pr \{ E_1^r \} + \sum_{j=2}^4 \Pr \{ E_j^r | (E_1^r)^c \}, \quad (\text{A.1})$$

where the inequality in (A.1) follows from the union bound. From the asymptotic equipartition property (AEP), [36, ch. 5.1], it follows that for sufficiently large n , $\Pr \{ E_1^r \}$ can be bounded by ϵ .

The event E_2^r is the union of the following events

$$E_2^r = \bigcup_{\tilde{u}_1 \in \{1, 2, \dots, 2^{n\hat{R}_1^r}\}, \tilde{u}_1 \neq u_{1,b}^r} E_2^r(\tilde{u}_1), \quad (\text{A.2})$$

where $E_2^r(\tilde{u}_1)$ is defined as follows

$$E_2^r(\tilde{u}_1) \triangleq \{ (\mathbf{x}_1(\tilde{u}_1, u_{1,b-1}^d), \mathbf{x}_2(u_{2,b}^r, u_{2,b-1}^d), \mathbf{v}_1(u_{1,b-1}^d), \mathbf{v}_2(u_{2,b-1}^d), \mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d), \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)} | \tilde{u}_1 \neq u_{1,b}^r \}. \quad (\text{A.3})$$

Let us denote $\mathbf{x}_1(\tilde{u}_1, u_{1,b-1}^d)$ by $\tilde{\mathbf{x}}_1$, $\mathbf{x}_2(u_{2,b}^r, u_{2,b-1}^d)$ by \mathbf{x}_2 , $\mathbf{v}_1(u_{1,b-1}^d)$ by \mathbf{v}_1 , $\mathbf{v}_2(u_{2,b-1}^d)$ by \mathbf{v}_2 , $\mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d)$ by \mathbf{x}_3 and $\mathbf{y}_{3,b}$ by \mathbf{y}_3 . The joint distribution of $\tilde{\mathbf{x}}_1$, \mathbf{x}_2 , \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{x}_3 and \mathbf{y}_3 obeys

$$p(\tilde{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3, \mathbf{y}_3) = \prod_{j=1}^n p(y_{3,j} | x_{2,j}, x_{3,j}, v_{1,j}, v_{2,j}, x_{3,j}) p(x_{2,j}, v_{1,j}, v_{2,j}, x_{3,j}) p(\tilde{x}_{1,j} | v_{1,j}). \quad (\text{A.4})$$

Using the chain (A.4) in [35, Thm. 14.2.3], $\Pr \{ E_2^r(\tilde{u}_1) | (E_1^r)^c \}$ can be bounded by

$$\begin{aligned} \Pr \{ E_2^r(\tilde{u}_1) | (E_1^r)^c \} &\leq 2^{-n[I(X_1; Y_3 | X_2, V_1, V_2, X_3) - \epsilon_1]}, \\ &\stackrel{(a)}{=} 2^{-n[I(X_1; Y_3 | X_2, V_1, X_3) - \epsilon_1]}, \end{aligned} \quad (\text{A.5})$$

where (a) follows from the Markov chain $V_2 - (X_2, X_3) - (V_1, X_1, Y_3)$. Hence,

$$\begin{aligned} \Pr \{ E_2^r | (E_1^r)^c \} &\leq \sum_{\substack{\tilde{u}_1 \in \{1, 2, \dots, 2^{n\hat{R}_1^r}\} \\ \tilde{u}_1 \neq u_{1,b}^r}} \Pr \{ E_2^r(\tilde{u}_1) | (E_1^r)^c \} \\ &\leq 2^{n[\hat{R}_1^r - (I(X_1; Y_3 | X_2, V_1, X_3) - \epsilon_1)]}, \end{aligned} \quad (\text{A.6})$$

which can be bounded by ϵ , for large enough n , if

$$\hat{R}_1^r < I(X_1; Y_3 | X_2, V_1, X_3) - \epsilon_1. \quad (\text{A.7})$$

By following similar arguments to these in (A.2)–(A.7), we can also show that $\Pr \{ E_3^r | (E_1^r)^c \}$ can be bounded by ϵ , for large enough n , if

$$\hat{R}_2^r < I(X_2; Y_3 | X_1, V_2, X_3) - \epsilon_1. \quad (\text{A.8})$$

²The overall average probability of channel decoding error at the relay calculated over all B blocks is upper bounded by $\sum_{b=1}^B \bar{P}_{r,ch}^{(n,b)}$, where $\bar{P}_{r,ch}^{(n,b)}$ is the average probability of error at block b assuming correct decoding at block $b-1$. We show that $\bar{P}_{r,ch}^{(n,b)}$ is upper bounded by a quantity independent of b , denoted $\bar{P}_{r,ch}^{(n)}$, i.e. $\bar{P}_{r,ch}^{(n,b)} \leq \bar{P}_{r,ch}^{(n)}$. Thus, the average probability of error is upper bounded by $B \cdot \bar{P}_{r,ch}^{(n)}$, which goes to zero, for sufficiently large n , for any fixed B .

The event E_4^r is a union of the following events

$$E_4^r = \bigcup_{\tilde{u}_1 \in \{1, 2, \dots, 2^{n\hat{R}_1^r}\}, \tilde{u}_1 \neq u_{1,b}^r} \bigcup_{\tilde{u}_2 \in \{1, 2, \dots, 2^{n\hat{R}_2^r}\}, \tilde{u}_2 \neq u_{2,b}^r} E_4^r(\tilde{u}_1, \tilde{u}_2), \quad (\text{A.9})$$

where $E_4^r(\tilde{u}_1, \tilde{u}_2)$ is defined as follows

$$\begin{aligned} E_4^r(\tilde{u}_1, \tilde{u}_2) &\triangleq \{(\mathbf{x}_1(\tilde{u}_1, u_{1,b-1}^d), \mathbf{x}_2(\tilde{u}_2, u_{2,b-1}^d), \mathbf{v}_1(u_{1,b-1}^d), \mathbf{v}_2(u_{2,b-1}^d), \mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d), \mathbf{y}_{3,b}) \\ &\in A_\epsilon^{*(n)}(X_1, X_2, V_1, V_2, X_3, Y_3) | \tilde{u}_1 \neq u_{1,b}^r, \tilde{u}_2 \neq u_{2,b}^r\}. \end{aligned} \quad (\text{A.10})$$

Let us denote $\mathbf{x}_1(\tilde{u}_1, u_{1,b-1}^d)$ by $\tilde{\mathbf{x}}_1$, $\mathbf{x}_2(\tilde{u}_2, u_{2,b-1}^d)$ by $\tilde{\mathbf{x}}_2$, $\mathbf{v}_1(u_{1,b-1}^d)$ by \mathbf{v}_1 , $\mathbf{v}_2(u_{2,b-1}^d)$ by \mathbf{v}_2 , $\mathbf{x}_3(u_{1,b-1}^d, u_{2,b-1}^d)$ by \mathbf{x}_3 and $\mathbf{y}_{3,b}$ by \mathbf{y}_3 . The joint distribution of $\tilde{\mathbf{x}}_1$, $\tilde{\mathbf{x}}_2$, \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{x}_3 and \mathbf{y}_3 obeys

$$p(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3, \mathbf{y}_3) = \prod_{j=1}^n p(y_{3,j} | v_{1,j}, v_{2,j}, x_{3,j}) p(v_{1,j}, v_{2,j}, x_{3,j}) p(\tilde{x}_{1,j}, \tilde{x}_{2,j} | v_{1,j}, v_{2,j}), \quad (\text{A.11})$$

Using the chain (A.11) in [35, Thm. 14.2.3], $\Pr\{E_4^r(\tilde{u}_1, \tilde{u}_2) | (E_1^r)^c\}$ can be bounded by

$$\Pr\{E_4^r(\tilde{u}_1, \tilde{u}_2) | (E_1^r)^c\} \leq 2^{-n[I(X_1, X_2; Y_3 | V_1, V_2, X_3) - \epsilon_1]}. \quad (\text{A.12})$$

Hence,

$$\begin{aligned} \Pr\{E_4^r | (E_1^r)^c\} &\leq \sum_{\tilde{u}_1 \in \{1, 2, \dots, 2^{n\hat{R}_1^r}\}, \tilde{u}_1 \neq u_{1,b}^r} \sum_{\tilde{u}_2 \in \{1, 2, \dots, 2^{n\hat{R}_2^r}\}, \tilde{u}_2 \neq u_{2,b}^r} \Pr\{E_4^r(\tilde{u}_1, \tilde{u}_2) | (E_1^r)^c\} \\ &\leq 2^{n[\hat{R}_1^r + \hat{R}_2^r - (I(X_1, X_2; Y_3 | V_1, V_2, X_3) - \epsilon_1)]}, \end{aligned} \quad (\text{A.13})$$

which can be bounded by ϵ , for large enough n , if

$$\hat{R}_1^r + \hat{R}_2^r < I(X_1, X_2; Y_3 | V_1, V_2, X_3) - \epsilon_1. \quad (\text{A.14})$$

Hence if conditions (A.7), (A.8) and (A.14) hold, for large enough n , $\bar{P}_{r, ch}^{(n)} \leq 4\epsilon$.

Relay source decoder: A relay source decoding error event, E_{src}^r , in block b , is the union of the following events:

$$\begin{aligned} E_5^r &\triangleq \{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_{3,b}) \notin A_\epsilon^{*(m)}(S_1, S_2, W_3)\}, \\ E_6^r &\triangleq \{\exists \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b} : f_1^r(\tilde{\mathbf{s}}_1) = f_1^r(\mathbf{s}_{1,b}) \text{ and } (\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b}, \mathbf{w}_{3,b}) \in A_\epsilon^{*(m)}(S_1, S_2, W_3)\}, \\ E_7^r &\triangleq \{\exists \tilde{\mathbf{s}}_2 \neq \mathbf{s}_{2,b} : f_2^r(\tilde{\mathbf{s}}_2) = f_2^r(\mathbf{s}_{2,b}) \text{ and } (\mathbf{s}_{1,b}, \tilde{\mathbf{s}}_2, \mathbf{w}_{3,b}) \in A_\epsilon^{*(m)}(S_1, S_2, W_3)\}, \\ E_8^r &\triangleq \{\exists \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b}, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_{2,b} : f_1^r(\tilde{\mathbf{s}}_1) = f_1^r(\mathbf{s}_{1,b}), f_2^r(\tilde{\mathbf{s}}_2) = f_2^r(\mathbf{s}_{2,b}) \text{ and } (\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{w}_{3,b}) \in A_\epsilon^{*(m)}(S_1, S_2, W_3)\}. \end{aligned}$$

The average probability of source decoding error at the relay in block b , $\bar{P}_{r, src}^{(m)}$, is defined by

$$\bar{P}_{r, src}^{(m)} = \Pr\left\{E_5^r \bigcup \bigcup_{j=6}^8 \{E_j^r \cap (E_5^r)^c\}\right\} \leq \Pr\{E_5^r\} + \sum_{j=6}^8 \Pr\{E_j^r | (E_5^r)^c\} \quad (\text{A.15})$$

where the inequality in (A.15) follows from the union bound. From the AEP for sufficiently large n , $\Pr\{E_5^r\}$ can be bounded by ϵ .

To bound $\Pr \{E_6^r | (E_5^r)^c\}$ we have

$$\begin{aligned}
\Pr \{E_6^r | (E_5^r)^c\} &= \sum_{(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3) \in A_\epsilon^{*(m)}(S_1, S_2, W_3)} p(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3) \times \\
&\quad \Pr \left\{ \exists \tilde{\mathbf{s}}_1 \neq \mathbf{s}_1 : f_1^r(\tilde{\mathbf{s}}_1) = f_1^r(\mathbf{s}_1), (\tilde{\mathbf{s}}_1, \mathbf{s}_2, \mathbf{w}_3) \in A_\epsilon^{*(m)}(S_1, S_2, W_3) | \mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3 \right\} \\
&\leq \sum_{(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3) \in A_\epsilon^{*(m)}(S_1, S_2, W_3)} p(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3) \left(\sum_{\substack{\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1 \\ \tilde{\mathbf{s}}_1 \in A_\epsilon^{*(m)}(S_1 | \mathbf{s}_2, \mathbf{w}_3)}} \Pr \{f_1^r(\tilde{\mathbf{s}}_1) = f_1^r(\mathbf{s}_1) | \mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3\} \right) \\
&\stackrel{(a)}{\leq} \sum_{(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3) \in A_\epsilon^{*(m)}(S_1, S_2, W_3)} p(\mathbf{s}_1, \mathbf{s}_2, \mathbf{w}_3) 2^{-mR_1^r} \left| A_\epsilon^{*(m)}(S_1 | \mathbf{s}_2, \mathbf{w}_3) \right| \\
&\stackrel{(b)}{\leq} 2^{-m[R_1^r - H(S_1 | S_2, W_3) - \epsilon_1]}, \tag{A.16}
\end{aligned}$$

where (a) follows as the bin assignments are independent and uniform; and (b) follows from [35, Thm. 14.2.2].

The RHS of inequality (A.16) can be bounded by ϵ , for large enough m , if

$$H(S_1 | S_2, W_3) + \epsilon_1 < R_1^r. \tag{A.17}$$

Following similar arguments as in (A.16), we can also show that $\Pr \{E_7^r | (E_5^r)^c\}$ can be bounded by ϵ , for large enough m , if

$$H(S_2 | S_1, W_3) + \epsilon_1 < R_2^r. \tag{A.18}$$

and $\Pr \{E_8^r | (E_5^r)^c\}$ can be bounded by ϵ , for large enough m , if

$$H(S_1, S_2 | W_3) + \epsilon_1 < R_1^r + R_2^r. \tag{A.19}$$

Hence, if conditions (A.17), (A.18) and (A.19) hold, for large enough m , $\bar{P}_{r,src}^{(m)} \leq 4\epsilon$. Combining conditions (A.7), (A.8) and (A.14) with conditions (A.17), (A.18) and (A.19) and using $\hat{R}_i^r = \frac{1}{\kappa} R_i^r, i = 1, 2$ yields the relay decoding constrains (8a)–(8c) in Thm. 1, for the achievability of rate κ .

Next the destination error probability analysis is derived.

Destination channel decoder: Assuming correct decoding at block $b + 1$, a destination channel decoding error event, E_{ch}^d , in block b , is the union of the following events:

$$\begin{aligned}
E_1^d &\triangleq \{(\mathbf{x}_1(u_{1,b+1}^r, u_{1,b}^d), \mathbf{x}_2(u_{2,b+1}^r, u_{2,b}^d), \mathbf{v}_1(u_{1,b}^d), \mathbf{v}_2(u_{2,b}^d), \mathbf{x}_3(u_{1,b}^d, u_{2,b}^d), \mathbf{y}_{b+1}) \notin A_\epsilon^{*(n)}\}, \\
E_2^d &\triangleq \{\exists \tilde{u}_1 \neq u_{1,b}^d : (\mathbf{x}_1(u_{1,b+1}^r, \tilde{u}_1), \mathbf{x}_2(u_{2,b+1}^r, u_{2,b}^d), \mathbf{v}_1(\tilde{u}_1), \mathbf{v}_2(u_{2,b}^d), \mathbf{x}_3(\tilde{u}_1, u_{2,b}^d), \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)}\}, \\
E_3^d &\triangleq \{\exists \tilde{u}_2 \neq u_{2,b}^d : (\mathbf{x}_1(u_{1,b+1}^r, u_{1,b}^d), \mathbf{x}_2(u_{2,b+1}^r, \tilde{u}_2), \mathbf{v}_1(u_{1,b}^d), \mathbf{v}_2(\tilde{u}_2), \mathbf{x}_3(u_{1,b}^d, \tilde{u}_2), \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)}\}, \\
E_4^d &\triangleq \{\exists \tilde{u}_1 \neq u_{1,b}^d, \tilde{u}_2 \neq u_{2,b}^d : (\mathbf{x}_1(u_{1,b+1}^r, \tilde{u}_1), \mathbf{x}_2(u_{2,b+1}^r, \tilde{u}_2), \mathbf{v}_1(\tilde{u}_1), \mathbf{v}_2(\tilde{u}_2), \mathbf{x}_3(\tilde{u}_1, \tilde{u}_2), \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)}\}.
\end{aligned}$$

The average probability of channel decoding error at the destination in block b , $\bar{P}_{d,ch}^{(n)}$, is defined by

$$\bar{P}_{d,ch}^{(n)} = \Pr \left\{ E_1^d \bigcup \bigcup_{j=2}^4 \{E_j^d \cap (E_1^d)^c\} \right\} \leq \Pr \{E_1^d\} + \sum_{j=2}^4 \Pr \{E_j^d | (E_1^d)^c\}, \tag{A.20}$$

where the inequality in (A.20) follows from the union bound. From the AEP it follows that for sufficiently large n , $\Pr \{E_1^d\}$ can be bounded by ϵ . The event E_2^d is a union of the following events

$$E_2^d = \bigcup_{\tilde{u}_1 \in \{1, 2, \dots, 2^{n\hat{R}_1^d}\}, \tilde{u}_1 \neq u_{1,b}^d} E_2^d(\tilde{u}_1) \quad (\text{A.21})$$

where $E_2^d(\tilde{u}_1)$ is defined as follows

$$\begin{aligned} E_2^d(\tilde{u}_1) \triangleq & \{(\mathbf{x}_1(u_{1,b+1}^r, \tilde{u}_1), \mathbf{x}_2(u_{2,b+1}^r, u_{2,b}^d), \mathbf{v}_1(\tilde{u}_1), \mathbf{v}_2(u_{2,b}^d), \mathbf{x}_3(\tilde{u}_1, u_{2,b}^d), \mathbf{y}_{b+1}) \\ & \in A_\epsilon^{*(n)}(X_1, X_2, V_1, V_2, X_3, Y) \mid \tilde{u}_1 \neq u_{1,b}^d\}. \end{aligned} \quad (\text{A.22})$$

Using [35, Thm. 14.2.3], $\Pr \{E_2^d(\tilde{u}_1) | (E_1^d)^c\}$ can be bounded by

$$\begin{aligned} \Pr \{E_2^d(\tilde{u}_1) | (E_1^d)^c\} & \leq 2^{-n[I(X_1, X_3, V_1; Y | X_2, V_2) - \epsilon_1]}, \\ & \stackrel{(a)}{=} 2^{-n[I(X_1, X_3; Y | X_2, V_2) - \epsilon_1]}, \end{aligned} \quad (\text{A.23})$$

where (a) follows from the Markov chain $V_1 - (X_1, X_3) - Y$. Hence,

$$\begin{aligned} \Pr \{E_2^d | (E_1^d)^c\} & \leq \sum_{\tilde{u}_1 \in \{1, 2, \dots, 2^{n\hat{R}_1^d}\}, \tilde{u}_1 \neq u_{1,b}^d} \Pr \{E_2^d(\tilde{u}_1) | (E_1^d)^c\} \\ & \leq 2^{n[\hat{R}_1^d - (I(X_1, X_3; Y | X_2, V_2) - \epsilon_1)]}, \end{aligned} \quad (\text{A.24})$$

which can be bounded by ϵ , for large enough n , if

$$\hat{R}_1^d < I(X_1, X_3; Y | X_2, V_2) - \epsilon_1. \quad (\text{A.25})$$

Following similar arguments as in (A.21)–(A.25), we can also show that $\Pr \{E_3^d | (E_1^d)^c\}$ can be bounded by ϵ , for large enough n , if

$$\hat{R}_2^d < I(X_2, X_3; Y | X_1, V_1) - \epsilon_1, \quad (\text{A.26})$$

and $\Pr \{E_4^d | (E_1^d)^c\}$ can be bounded by ϵ , for large enough n , if

$$\hat{R}_1^d + \hat{R}_2^d < I(X_1, X_2, X_3; Y_3) - \epsilon_1. \quad (\text{A.27})$$

Hence, if conditions (A.25), (A.26) and (A.27) hold, for large enough n , $\bar{P}_{d, ch}^{(n)} \leq 4\epsilon$.

Destination source decoder: The destination source decoding error probability analysis is similar to the relay source decoding error probability analysis. The average probability of source decoder error at the destination in block b (defined in the same manner as in the relay) can be bounded by ϵ , for sufficiently large m , if

$$H(S_1 | S_2, W) + \epsilon_1 < R_1^d, \quad (\text{A.28a})$$

$$H(S_2 | S_1, W) + \epsilon_1 < R_2^d, \quad (\text{A.28b})$$

$$H(S_1, S_2 | W) + \epsilon_1 < R_1^d + R_2^d. \quad (\text{A.28c})$$

Combining conditions (A.25), (A.26) and (A.27) with conditions (A.28) and using $\hat{R}_i^d = \frac{1}{\kappa} R_i^d, i = 1, 2$ yields the destination decoding constraints (8d)–(8f) in Thm. 1, for the achievability of rate κ .

This concludes the error probability analysis for Thm. 1.

APPENDIX B

PROOF OF THM. 3

Let $P_e^{(m,n)} \rightarrow 0$ as $n, m \rightarrow \infty$, for a sequence of encoders and decoders $f_1^{(m,n)}, f_2^{(m,n)}, f_3^{(m,n)}, g^{(m,n)}, g_3^{(m,n)}$, such that $\kappa = n/m$ is fixed. We will use Fano's inequality which states

$$H(S_1^m, S_2^m | \hat{S}_{1,3}^m, \hat{S}_{2,3}^m) \leq 1 + mP_e^{(m,n)} \log |\mathcal{S}_1 \times \mathcal{S}_2|$$

$$\triangleq m\delta(P_e^{(m,n)}), \quad (\text{B.1})$$

where $\delta(x)$ is a non-negative function that approaches $\frac{1}{m}$ as $x \rightarrow 0$. We also obtain

$$H(S_1^m, S_2^m | \hat{S}_{1,3}^m, \hat{S}_{2,3}^m) \geq H(S_1^m, S_2^m | \hat{S}_{1,3}^m, \hat{S}_{2,3}^m, Y_3^n, W_3^m) \quad (\text{B.2a})$$

$$= H(S_1^m, S_2^m | Y_3^n, W_3^m) \quad (\text{B.2b})$$

$$\geq H(S_1^m | Y_3^n, W_3^m, S_2^m), \quad (\text{B.2c})$$

where (B.2a) follows from the fact that conditioning reduces entropy; (B.2b) follow from the fact that $(\hat{S}_{1,3}^m, \hat{S}_{2,3}^m)$ is a deterministic function of Y_3^n and W_3^m ; and (B.2c) follows from non-negativity of the entropy function for discrete sources. Constraint (14a) is a consequence of the following chain of inequalities:

$$\sum_{k=1}^n I(X_{1,k}; Y_{3,k} | X_{2,k}, X_{3,k}) = \sum_{k=1}^n \left[H(Y_{3,k} | X_{2,k}, X_{3,k}) - H(Y_{3,k} | X_{1,k}, X_{2,k}, X_{3,k}) \right]$$

$$= \sum_{k=1}^n \left[H(Y_{3,k} | X_{2,k}, X_{3,k}) - H(Y_{3,k} | X_{1,1}^k, X_{2,1}^k, X_{3,1}^k, Y_{3,1}^{k-1}, Y^{k-1}) \right] \quad (\text{B.3a})$$

$$= \sum_{k=1}^n \left[H(Y_{3,k} | X_{2,k}, X_{3,k}) - H(Y_{3,k} | S_1^m, S_2^m, W_3^m, W^m, X_{1,1}^k, X_{2,1}^k, X_{3,1}^k, Y_{3,1}^{k-1}, Y^{k-1}) \right] \quad (\text{B.3b})$$

$$\geq \sum_{k=1}^n \left[H(Y_{3,k} | S_2^m, W_3^m, Y_{3,1}^{k-1}, X_{2,k}, X_{3,k}) - H(Y_{3,k} | S_1^m, S_2^m, W_3^m, W^m, Y_{3,1}^{k-1}) \right] \quad (\text{B.3c})$$

$$= \sum_{k=1}^n \left[H(Y_{3,k} | S_2^m, W_3^m, Y_{3,1}^{k-1}) - H(Y_{3,k} | S_1^m, S_2^m, W_3^m, W^m, Y_{3,1}^{k-1}) \right] \quad (\text{B.3d})$$

$$= \sum_{k=1}^n I(S_1^m, W^m; Y_{3,k} | Y_{3,1}^{k-1}, S_2^m, W_3^m)$$

$$= I(S_1^m, W^m; Y_3^n | S_2^m, W_3^m) \quad (\text{B.3e})$$

$$\geq I(S_1^m; Y_3^n | S_2^m, W_3^m) \quad (\text{B.3f})$$

$$= H(S_1^m | S_2^m, W_3^m) - H(S_1^m | Y_3^n, S_2^m, W_3^m)$$

$$\geq mH(S_1 | S_2, W_3) - m\delta(P_e^{(m,n)}), \quad (\text{B.3g})$$

where (B.3a) follows from the memoryless channel assumption; (B.3b) follows from the Markov relation $(S_1^m, S_2^m, W_3^m, W^m) - (X_{1,1}^k, X_{2,1}^k, X_{3,1}^k, Y_{3,1}^{k-1}, Y^{k-1}) - Y_{3,k}$ (see [40]); (B.3c) follows from the fact that conditioning reduces entropy; (B.3d) follows from the fact that $X_{2,k}$ is a deterministic function of S_2^m , and $X_{3,k}$ is a deterministic function of $(W_3^m, Y_{3,1}^{k-1})$; (B.3e) follows from the chain rule for mutual information; (B.3f) follows from the non-negativity of the mutual information; and (B.3g) follows from the memoryless sources and side information assumption and from equations (B.1)–(B.2).

Following arguments similar to those that led to (B.3g) we can also show that

$$\sum_{k=1}^n I(X_{2,k}; Y_{3,k} | X_{1,k}, X_{3,k}) \geq mH(S_2 | S_1, W_3) - m\delta(P_e^{(m,n)}) \quad (\text{B.4a})$$

$$\sum_{k=1}^n I(X_{1,k}, X_{2,k}; Y_{3,k} | X_{3,k}) \geq mH(S_1, S_2 | W_3) - m\delta(P_e^{(m,n)}). \quad (\text{B.4b})$$

From (B.3)–(B.4), we have that if κ is achievable for the MABRC then

$$\frac{1}{n} \sum_{k=1}^n I(X_{1,k}; Y_{3,k} | X_{2,k}, X_{3,k}) \geq \frac{1}{\kappa} (H(S_1 | S_2, W_3) - \epsilon), \quad (\text{B.5a})$$

$$\frac{1}{n} \sum_{k=1}^n I(X_{2,k}; Y_{3,k} | X_{1,k}, X_{3,k}) \geq \frac{1}{\kappa} (H(S_2 | S_1, W_3) - \epsilon), \quad (\text{B.5b})$$

$$\frac{1}{n} \sum_{k=1}^n I(X_{1,k}, X_{2,k}; Y_{3,k} | X_{3,k}) \geq \frac{1}{\kappa} (H(S_1, S_2 | W_3) - \epsilon). \quad (\text{B.5c})$$

for any $\epsilon > 0$ and large enough n . Using the concavity of mutual information over the set of all joint distributions $p(x_1, x_2, x_3)$, taking the limit as $m, n \rightarrow \infty$ and letting $P_e^{(m,n)} \rightarrow 0$ leads to conditions (14).

APPENDIX C

ERROR PROBABILITY ANALYSIS FOR THM. 7

We start with the relay error probability analysis. The average decoding error probability at the relay in block b , $\bar{P}_r^{(n)}$, is defined by

$$\begin{aligned} \bar{P}_r^{(n)} &\triangleq \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \in \mathcal{S}_1^n \times \mathcal{S}_2^n} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \Pr \{ \text{relay error} | (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \text{ are the source outputs} \} \\ &\leq \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \notin A_\epsilon^{*(n)}(S_1, S_2)} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) + \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \in A_\epsilon^{*(n)}(S_1, S_2)} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \Pr \{ \text{relay error} | (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \text{ are the source outputs} \} \end{aligned} \quad (\text{C.1a})$$

$$\begin{aligned} &\leq \epsilon + \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \in A_\epsilon^{*(n)}(S_1, S_2)} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \sum_{\mathbf{w}_{3,b} \in \mathcal{W}_3^n} p(\mathbf{w}_{3,b} | \mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \times \\ &\quad \Pr \{ \text{relay error} | (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \text{ are the source outputs, } \mathbf{w}_{3,b} \text{ is the side information} \} \end{aligned} \quad (\text{C.1b})$$

$$\begin{aligned} &\leq \epsilon + \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \in A_\epsilon^{*(n)}(S_1, S_2)} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \sum_{\mathbf{w}_{3,b} \notin A_\epsilon^{*(n)}(W_3 | \mathbf{s}_{1,b}, \mathbf{s}_{2,b})} p(\mathbf{w}_{3,b} | \mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \\ &\quad + \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \in A_\epsilon^{*(n)}(S_1, S_2)} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \sum_{\mathbf{w}_{3,b} \in A_\epsilon^{*(n)}(W_3 | \mathbf{s}_{1,b}, \mathbf{s}_{2,b})} p(\mathbf{w}_{3,b} | \mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \times \\ &\quad \Pr \{ \text{relay error} | (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \text{ are the source outputs, } \mathbf{w}_{3,b} \text{ is the side information} \} \end{aligned} \quad (\text{C.1c})$$

$$\begin{aligned} &\leq 2\epsilon + \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_{3,b}) \in A_\epsilon^{*(n)}(S_1, S_2, W_3)} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_{3,b}) \times \\ &\quad \Pr \{ \text{relay error} | (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \text{ are the source outputs, } \mathbf{w}_{3,b} \text{ is the side information} \}, \end{aligned} \quad (\text{C.1d})$$

where (C.1a), (C.1c) follows from the union bound and (C.1b), (C.1d) follows from the AEP, for sufficiently large n .

In the following we show that for $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_{3,b}) \in A_\epsilon^{*(n)}(S_1, S_2, W_3)$, the summation terms in (C.1d) can be upper bounded independently of $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_{3,b})$. In order to show the above we assume that $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_{3,b}) \in A_\epsilon^{*(n)}(S_1, S_2, W_3)$ and let \mathcal{D} denote the event that the triplet $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_{3,b})$ is the sources outputs and the side information at the relay.

Let ϵ_0 be a positive number such that $\epsilon_0 > \epsilon$ and $\epsilon_0 \rightarrow 0$ as $\epsilon \rightarrow 0$. Assuming correct decoding at block $b-1$ (hence $(u_{1,b-1}, u_{2,b-1})$ are available at the relay), a relay decoding error event, E^r , in block b , is the union of the following events:

$$\begin{aligned} E_1^r &\triangleq \{ (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{x}_1(\mathbf{s}_{1,b}, u_{1,b-1}), \mathbf{x}_2(\mathbf{s}_{2,b}, u_{2,b-1}), \mathbf{v}_1(u_{1,b-1}), \mathbf{v}_2(u_{2,b-1}), \mathbf{x}_3(u_{1,b-1}, u_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \notin A_\epsilon^{*(n)} \}, \\ E_2^r &\triangleq \{ \exists (\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \neq (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) : \\ &\quad (\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_1(\tilde{\mathbf{s}}_1, u_{1,b-1}), \mathbf{x}_2(\tilde{\mathbf{s}}_2, u_{2,b-1}), \mathbf{v}_1(u_{1,b-1}), \mathbf{v}_2(u_{2,b-1}), \mathbf{x}_3(u_{1,b-1}, u_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)} \}. \end{aligned}$$

From the AEP, for sufficiently large n , $\Pr \{ E_1^r | \mathcal{D} \}$ can be bounded by ϵ . Therefore by the union bound

$$\Pr \{ E^r | \mathcal{D} \} \leq \epsilon + \Pr \{ E_2^r | \mathcal{D}, (E_1^r)^c \}. \quad (\text{C.2})$$

The event E_2^r is the union of the following events:

$$\begin{aligned}
E_{21}^r &\triangleq \{ \exists \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b} : \\
&\quad (\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b}, \mathbf{x}_1(\tilde{\mathbf{s}}_1, u_{1,b-1}), \mathbf{x}_2(\mathbf{s}_{2,b}, u_{2,b-1}), \mathbf{v}_1(u_{1,b-1}), \mathbf{v}_2(u_{2,b-1}), \mathbf{x}_3(u_{1,b-1}, u_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)} \}, \\
E_{22}^r &\triangleq \{ \exists \tilde{\mathbf{s}}_2 \neq \mathbf{s}_{2,b} : \\
&\quad (\mathbf{s}_{1,b}, \tilde{\mathbf{s}}_2, \mathbf{x}_1(\mathbf{s}_{1,b}, u_{1,b-1}), \mathbf{x}_2(\tilde{\mathbf{s}}_2, u_{2,b-1}), \mathbf{v}_1(u_{1,b-1}), \mathbf{v}_2(u_{2,b-1}), \mathbf{x}_3(u_{1,b-1}, u_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)} \}, \\
E_{23}^r &\triangleq \{ \exists \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b}, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_{2,b} : \\
&\quad (\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_1(\tilde{\mathbf{s}}_1, u_{1,b-1}), \mathbf{x}_2(\tilde{\mathbf{s}}_2, u_{2,b-1}), \mathbf{v}_1(u_{1,b-1}), \mathbf{v}_2(u_{2,b-1}), \mathbf{x}_3(u_{1,b-1}, u_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)} \}.
\end{aligned}$$

Hence, by the union bound

$$\Pr \{ E_2^r | \mathcal{D}, (E_1^r)^c \} \leq \sum_{j=1}^3 \Pr \{ E_{2j}^r | \mathcal{D}, (E_1^r)^c \}. \quad (\text{C.3})$$

To bound $\Pr \{ E_{21}^r | \mathcal{D}, (E_1^r)^c \}$ we define the event $E_{21}^r(\tilde{\mathbf{s}}_1)$ as follows

$$\begin{aligned}
E_{21}^r(\tilde{\mathbf{s}}_1) &\triangleq \{ (\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b}, \mathbf{x}_1(\tilde{\mathbf{s}}_1, u_{1,b-1}), \mathbf{x}_2(\mathbf{s}_{2,b}, u_{2,b-1}), \mathbf{v}_1(u_{1,b-1}), \mathbf{v}_2(u_{2,b-1}), \\
&\quad \mathbf{x}_3(u_{1,b-1}, u_{2,b-1}), \mathbf{w}_{3,b}, \mathbf{y}_{3,b}) \in A_\epsilon^{*(n)}(S_1, S_2, X_1, X_2, V_1, V_2, X_3, W_3, Y_3) | \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b} \}. \quad (\text{C.4})
\end{aligned}$$

Let us denote $\mathbf{s}_{2,b}$ by \mathbf{s}_2 , $\mathbf{x}_1(\tilde{\mathbf{s}}_1, u_{1,b-1})$ by $\tilde{\mathbf{x}}_1$, $\mathbf{x}_2(\mathbf{s}_{2,b}, u_{2,b-1})$ by \mathbf{x}_2 , $\mathbf{v}_1(u_{1,b-1})$ by \mathbf{v}_1 , $\mathbf{v}_2(u_{2,b-1})$ by \mathbf{v}_2 , $\mathbf{x}_3(u_{1,b-1}, u_{2,b-1})$ by \mathbf{x}_3 , $\mathbf{w}_{3,b}$ by \mathbf{w}_3 and $\mathbf{y}_{3,b}$ by \mathbf{y}_3 . Note that the joint distribution of $\tilde{\mathbf{s}}_1, \mathbf{s}_2, \tilde{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3, \mathbf{w}_3$ and \mathbf{y}_3 obeys

$$\begin{aligned}
&p(\tilde{\mathbf{s}}_1, \mathbf{s}_2, \tilde{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3, \mathbf{w}_3, \mathbf{y}_3) \\
&= \prod_{j=1}^n p(\tilde{s}_{1,j}) p(s_{2,j}, w_{3,j}) p(x_{2,j}, y_{3,j} | s_{2,j}, w_{3,j}, v_{1,j}, v_{2,j}, x_{3,j}) \times \\
&\quad p(v_{1,j}, v_{2,j}, x_{3,j} | \tilde{s}_{1,j}, s_{2,j}, w_{3,j}) p(\tilde{x}_{1,j} | \tilde{s}_{1,j}, v_{1,j}, v_{2,j}, x_{3,j}). \quad (\text{C.5})
\end{aligned}$$

Equation (C.5) shows that the assumption of [9, Lemma, Appendix A] on the conditional distribution is satisfied.

Hence, we can use this lemma to bound $\Pr \{ E_{21}^r(\tilde{\mathbf{s}}_1) | \mathcal{D}, (E_1^r)^c \}$ with the following assignments:

$$\mathbf{z}_1 = (\mathbf{s}_2, \mathbf{w}_3), \mathbf{z}_2 = \tilde{\mathbf{s}}_1, \mathbf{Z}_3 = (\mathbf{X}_3, \mathbf{V}_1, \mathbf{V}_2), \mathbf{Z}_4 = \tilde{\mathbf{X}}_1, \mathbf{Z}_5 = (\mathbf{X}_2, \mathbf{Y}_3), \quad (\text{C.6})$$

as follows

$$\begin{aligned}
\Pr \{ E_{21}^r(\tilde{\mathbf{s}}_1) | \mathcal{D}, (E_1^r)^c \} &\leq 2^{-n[I(X_1; Y_3 | S_2, X_2, V_1, V_2, X_3, W_3) - \epsilon_0]} \\
&\stackrel{(a)}{=} 2^{-n[I(X_1; Y_3 | S_2, X_2, V_1, X_3, W_3) - \epsilon_0]}, \quad (\text{C.7})
\end{aligned}$$

where (a) follows from the Markov chain $V_2 - (S_2, X_2, V_1, X_3, W_3) - Y_3$. Hence

$$\begin{aligned}
\Pr \{E_{21}^r | \mathcal{D}, (E_1^r)^c\} &\leq \sum_{\substack{\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1 \\ \tilde{\mathbf{s}}_1 \in A_\epsilon^{*(n)}(S_1 | \mathbf{s}_{2,b}, \mathbf{w}_{3,b})}} \Pr \{E_{21}^r(\tilde{\mathbf{s}}_1) | \mathcal{D}, (E_1^r)^c\} \\
&\leq \sum_{\substack{\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1 \\ \tilde{\mathbf{s}}_1 \in A_\epsilon^{*(n)}(S_1 | \mathbf{s}_{2,b}, \mathbf{w}_{3,b})}} 2^{-n[I(X_1; Y_3 | S_2, X_2, V_1, X_3, W_3) - \epsilon_0]} \\
&\leq \left| A_\epsilon^{*(n)}(S_1 | \mathbf{s}_{2,b}, \mathbf{w}_{3,b}) \right| \cdot 2^{-n[I(X_1; Y_3 | S_2, X_2, V_1, X_3, W_3) - \epsilon_0]} \\
&\leq 2^{n[H(S_1 | S_2, W_3) - I(X_1; Y_3 | S_2, X_2, V_1, X_3, W_3) + 2\epsilon_0]}, \tag{C.8}
\end{aligned}$$

which can be bounded by ϵ , for large enough n , if

$$H(S_1 | S_2, W_3) < I(X_1; Y_3 | S_2, X_2, V_1, X_3, W_3) - 2\epsilon_0, \tag{C.9}$$

Following similar arguments as in (C.4)–(C.8), we can also show that $\Pr \{E_{22}^r | \mathcal{D}, (E_1^r)^c\}$ can be bounded by ϵ , for large enough n , if

$$H(S_2 | S_1, W_3) < I(X_2; Y_3 | S_1, X_1, V_2, X_3, W_3) - 2\epsilon_0, \tag{C.10}$$

and $\Pr \{E_{23}^r | \mathcal{D}, (E_1^r)^c\}$ can be bounded by ϵ , for large enough n , if

$$H(S_1, S_2 | W_3) < I(X_1, X_2; Y_3 | V_1, V_2, X_3, W_3) - 2\epsilon_0. \tag{C.11}$$

Hence, if conditions (40a)–(40c) hold, for large enough n ,

$$\Pr \{E_2^r | \mathcal{D}, (E_1^r)^c\} \leq \sum_{j=1}^3 \Pr \{E_{2j}^r | \mathcal{D}, (E_1^r)^c\} \leq 3\epsilon. \tag{C.12}$$

Combining equations (C.1), (C.2) and (C.12) yields

$$\bar{P}_r^{(n)} \leq \Pr \{E_2^r | \mathcal{D}, (E_1^r)^c\} + 3\epsilon \leq 6\epsilon. \tag{C.13}$$

Next the destination error probability analysis is derived.

Destination channel decoder: Let $(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1})$ be the output of the source in block $b+1$. Assuming correct decoding at block $b+1$, a destination channel decoding error event, E_{ch}^d , in block b , is the union of the following events:

$$E_1^d \triangleq \{(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1}, \mathbf{x}_1(\mathbf{s}_{1,b+1}, u_{1,b}), \mathbf{x}_2(\mathbf{s}_{2,b+1}, u_{2,b}), \mathbf{v}_1(u_{1,b}), \mathbf{v}_2(u_{2,b}), \mathbf{x}_3(u_{1,b}, u_{2,b}), \mathbf{w}_{b+1}, \mathbf{y}_{b+1}) \notin A_\epsilon^{*(n)}\},$$

$$E_2^d \triangleq \{\exists \tilde{u}_1 \neq u_{1,b} :$$

$$(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1}, \mathbf{x}_1(\mathbf{s}_{1,b+1}, \tilde{u}_1), \mathbf{x}_2(\mathbf{s}_{2,b+1}, u_{2,b}), \mathbf{v}_1(\tilde{u}_1), \mathbf{v}_2(u_{2,b}), \mathbf{x}_3(\tilde{u}_1, u_{2,b}), \mathbf{w}_{b+1}, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)}\},$$

$$E_3^d \triangleq \{\exists \tilde{u}_2 \neq u_{2,b} :$$

$$(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1}, \mathbf{x}_1(\mathbf{s}_{1,b+1}, u_{1,b}), \mathbf{x}_2(\mathbf{s}_{2,b+1}, \tilde{u}_2), \mathbf{v}_1(u_{1,b}), \mathbf{v}_2(\tilde{u}_2), \mathbf{x}_3(u_{1,b}, \tilde{u}_2), \mathbf{w}_{b+1}, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)}\},$$

$$E_4^d \triangleq \{\exists \tilde{u}_1 \neq u_{1,b}, \tilde{u}_2 \neq u_{2,b} :$$

$$(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1}, \mathbf{x}_1(\mathbf{s}_{1,b+1}, \tilde{u}_1), \mathbf{x}_2(\mathbf{s}_{2,b+1}, \tilde{u}_2), \mathbf{v}_1(\tilde{u}_1), \mathbf{v}_2(\tilde{u}_2), \mathbf{x}_3(\tilde{u}_1, \tilde{u}_2), \mathbf{w}_{b+1}, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)}\}.$$

The average probability of channel decoding error at the destination in block b , $\bar{P}_{d,ch}^{(n)}$, is then given by

$$\bar{P}_{d,ch}^{(n)} = \Pr \{E_1^d \bigcup \bigcup_{j=2}^4 E_j^d\} \leq \Pr \{E_1^d\} + \sum_{j=2}^4 \Pr \{E_j^d | (E_1^d)^c\}, \quad (\text{C.14})$$

where the inequality in (C.14) follows from the union bound. From the AEP for sufficiently large n , $\Pr\{E_1^d\}$ can be bounded by ϵ . The event E_2^d is the union of the following events

$$E_2^d = \bigcup_{\tilde{u}_1 \in \{1, 2, \dots, 2^{nR_1}\}, \tilde{u}_1 \neq u_{1,b}} E_2^d(\tilde{u}_1), \quad (\text{C.15})$$

where $E_2^d(\tilde{u}_1)$ is defined as follows

$$\begin{aligned} E_2^d(\tilde{u}_1) &\triangleq \{(\mathbf{s}_{1,b+1}, \mathbf{s}_{2,b+1}, \mathbf{x}_1(\mathbf{s}_{1,b+1}, \tilde{u}_1), \mathbf{x}_2(\mathbf{s}_{2,b+1}, u_{2,b}), \mathbf{v}_1(\tilde{u}_1), \mathbf{v}_2(u_{2,b}), \mathbf{x}_3(\tilde{u}_1, u_{2,b}), \mathbf{w}_{b+1}, \mathbf{y}_{b+1}) \\ &\in A_\epsilon^{*(n)}(S_1, S_2, X_1, X_2, V_1, V_2, X_3, W, Y) | \tilde{u}_1 \neq u_{1,b}\}. \end{aligned} \quad (\text{C.16})$$

Let us denote $\mathbf{s}_{1,b+1}$ by \mathbf{s}_1 , $\mathbf{s}_{2,b+1}$ by \mathbf{s}_2 , $\mathbf{x}_1(\mathbf{s}_{1,b+1}, \tilde{u}_1)$ by $\tilde{\mathbf{x}}_1$, $\mathbf{x}_2(\mathbf{s}_{2,b+1}, u_{2,b})$ by \mathbf{x}_2 , $\mathbf{v}_1(\tilde{u}_1)$ by $\tilde{\mathbf{v}}_1$, $\mathbf{v}_2(u_{2,b})$ by \mathbf{v}_2 , $\mathbf{x}_3(\tilde{u}_1, u_{2,b})$ by $\tilde{\mathbf{x}}_3$, \mathbf{w}_{b+1} by \mathbf{w} and \mathbf{y}_{b+1} by \mathbf{y} . The joint distribution of \mathbf{s}_1 , \mathbf{s}_2 , $\tilde{\mathbf{x}}_1$, \mathbf{x}_2 , $\tilde{\mathbf{v}}_1$, \mathbf{v}_2 , $\tilde{\mathbf{x}}_3$, \mathbf{w} and \mathbf{y} obeys

$$\begin{aligned} &p(\mathbf{s}_1, \mathbf{s}_2, \tilde{\mathbf{x}}_1, \mathbf{x}_2, \tilde{\mathbf{v}}_1, \mathbf{v}_2, \tilde{\mathbf{x}}_3, \mathbf{w}, \mathbf{y}) \\ &= \prod_{j=1}^n p(y_j | s_{1,j}, s_{2,j}, x_{2,j}, v_{2,j}, w_j) \times \\ &\quad p(\tilde{x}_{1,j}, \tilde{x}_{3,j}, \tilde{v}_{1,j} | s_{1,j}, s_{2,j}, x_{2,j}, v_{2,j}, w_j) p(s_{1,j}, s_{2,j}, x_{2,j}, v_{2,j}, w_j). \end{aligned} \quad (\text{C.17})$$

Using the chain (C.17) in [35, Thm. 14.2.3], $\Pr \{E_2^d(\tilde{u}_1) | (E_1^d)^c\}$ can be bounded by

$$\begin{aligned} \Pr \{E_2^d(\tilde{u}_1) | (E_1^d)^c\} &\leq 2^{-n[I(X_1, X_3, V_1; Y | S_1, S_2, X_2, V_2, W) - \epsilon_0]} \\ &\stackrel{(a)}{=} 2^{-n[I(X_1, X_3; Y | S_1, X_2, V_2) - \epsilon_0]}, \end{aligned} \quad (\text{C.18})$$

where (a) follows from the Markov chains $V_1 - (X_1, X_2, X_3, S_1, V_2, W) - Y$ and $(S_2, W) - (S_1, V_2, X_2) - Y$. Hence,

$$\begin{aligned} \Pr \{E_2^d | (E_1^d)^c\} &\leq \sum_{\tilde{u}_1 \in \{1, 2, \dots, 2^{nR_1}\}, \tilde{u}_1 \neq u_{1,b}} \Pr \{E_2^d(\tilde{u}_1) | (E_1^d)^c\} \\ &\leq 2^{n[R_1 - (I(X_1, X_3; Y | S_1, X_2, V_2) - \epsilon_0)]}, \end{aligned} \quad (\text{C.19})$$

which can be bounded by ϵ , for large enough n , if

$$R_1 < I(X_1, X_3; Y | S_1, X_2, V_2) - \epsilon_0. \quad (\text{C.20})$$

Following similar arguments as in (C.15)–(C.20), we can also show that $\Pr \{E_3^d | (E_1^d)^c\}$ can be bounded by ϵ , for large enough n , if

$$R_2 < I(X_2, X_3; Y | S_2, X_1, V_1) - \epsilon_0, \quad (\text{C.21})$$

and $\Pr \{E_4^d | (E_1^d)^c\}$ can be bounded by ϵ , for large enough n , if

$$R_1 + R_2 < I(X_1, X_2, X_3; Y | S_1, S_2) - \epsilon_0. \quad (\text{C.22})$$

Hence, if conditions (C.20), (C.21) and (C.22) hold, for large enough n , $\bar{P}_{d, ch}^{(n)} \leq 4\epsilon$.

Destination source decoder: The error probability analysis for decoding the sources at the destination is similar to the relay source decoder error probability analysis done for the separation-based achievable rate detailed in Section III. From this analysis it follows that the average probability of source decoding error at the destination in block b (defined in the same manner as in the relay, in Section III) can be made arbitrarily small, for sufficiently large n , if

$$H(S_1 | S_2, W) + \epsilon_0 < R_1, \quad (\text{C.23a})$$

$$H(S_2 | S_1, W) + \epsilon_0 < R_2, \quad (\text{C.23b})$$

$$H(S_1, S_2 | W) + \epsilon_0 < R_1 + R_2. \quad (\text{C.23c})$$

Combining conditions (C.20), (C.21) and (C.22) with conditions (C.23) yields the destination decoding constraints (40d)–(40f) in Thm. 7, for the achievability of rate $\kappa = 1$.

APPENDIX D

ERROR PROBABILITY ANALYSIS FOR THM. 8

We start with the relay error probability analysis.

Relay channel decoder: Assuming correct decoding at block $b - 1$, a relay channel decoding error event, E_{ch}^r , at block b , is the union of the following events:

$$\begin{aligned} E_1^r &\triangleq \{(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}, \mathbf{x}_1(u_{1,b}, \mathbf{s}_{1,b-1}), \mathbf{x}_2(u_{2,b}, \mathbf{s}_{2,b-1}), \mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}), \mathbf{w}_{3,b-1}, \mathbf{y}_b) \notin A_\epsilon^{*(n)}\}, \\ E_2^r &\triangleq \{\exists \tilde{u}_1 \neq u_{1,b} : (\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}, \mathbf{x}_1(\tilde{u}_{1,b}, \mathbf{s}_{1,b-1}), \mathbf{x}_2(u_{2,b}, \mathbf{s}_{2,b-1}), \mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}), \mathbf{w}_{3,b-1}, \mathbf{y}_b) \in A_\epsilon^{*(n)}\}, \\ E_3^r &\triangleq \{\exists \tilde{u}_2 \neq u_{2,b} : (\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}, \mathbf{x}_1(u_{1,b}, \mathbf{s}_{1,b-1}), \mathbf{x}_2(\tilde{u}_{2,b}, \mathbf{s}_{2,b-1}), \mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}), \mathbf{w}_{3,b-1}, \mathbf{y}_b) \in A_\epsilon^{*(n)}\}, \\ E_4^r &\triangleq \{\exists \tilde{u}_1 \neq u_{1,b}, \tilde{u}_2 \neq u_{2,b} : (\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}, \mathbf{x}_1(\tilde{u}_{1,b}, \mathbf{s}_{1,b-1}), \mathbf{x}_2(\tilde{u}_{2,b}, \mathbf{s}_{2,b-1}), \mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}), \mathbf{w}_{3,b-1}, \mathbf{y}_b) \in A_\epsilon^{*(n)}\}. \end{aligned}$$

The average probability of channel decoding error at the relay in block b , $\bar{P}_{r, ch}^{(n)}$, is defined by

$$\bar{P}_{r, ch}^{(n)} = \Pr \{E_1^r \bigcup \bigcup_{j=2}^4 E_j^r\} \leq \Pr \{E_1^r\} + \sum_{j=2}^4 \Pr \{E_j^r | (E_1^r)^c\}, \quad (\text{D.1})$$

where the inequality in (D.1) follows from the union bound. From the AEP for sufficiently large n , $\Pr \{E_1^r\}$ can be bounded by ϵ . Let ϵ_0 be a positive number such that $\epsilon_0 > \epsilon$ and $\epsilon_0 \rightarrow 0$ as $\epsilon \rightarrow 0$.

The event E_2^r is the union of the following events

$$E_2^r = \bigcup_{\tilde{u}_1 \in \{1, 2, \dots, 2^{nR_1}\}, \tilde{u}_1 \neq u_{1,b}} E_2^r(\tilde{u}_1), \quad (\text{D.2})$$

where $E_2^r(\tilde{u}_1)$ is defined as follows

$$E_2^r(\tilde{u}_1) : \left\{ (\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}, \mathbf{x}_1(\tilde{u}_1, \mathbf{s}_{1,b-1}), \mathbf{x}_2(u_{2,b}, \mathbf{s}_{2,b-1}), \mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1}), \mathbf{w}_{3,b-1}, \mathbf{y}_{3,b}) \right. \\ \left. \in A_\epsilon^{*(n)}(S_1, S_2, X_1, X_2, X_3, W_3, Y) \mid \tilde{u}_1 \neq u_{1,b} \right\}. \quad (\text{D.3})$$

Let us denote $\mathbf{s}_{1,b-1}$ by \mathbf{s}_1 , $\mathbf{s}_{2,b-1}$ by \mathbf{s}_2 , $\mathbf{x}_1(\tilde{u}_1, \mathbf{s}_{1,b-1})$ by $\tilde{\mathbf{x}}_1$, $\mathbf{x}_2(u_{2,b}, \mathbf{s}_{2,b-1})$ by \mathbf{x}_2 , $\mathbf{x}_3(\mathbf{s}_{1,b-1}, \mathbf{s}_{2,b-1})$ by \mathbf{x}_3 , $\mathbf{w}_{3,b-1}$ by \mathbf{w}_3 and $\mathbf{y}_{3,b}$ by \mathbf{y}_3 . The joint distribution of $\mathbf{s}_1, \mathbf{s}_2, \tilde{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{w}_3$ and \mathbf{y}_3 obeys

$$p(\mathbf{s}_1, \mathbf{s}_2, \tilde{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{w}_3, \mathbf{y}_3) \\ = \prod_{j=1}^n p(y_{3,j} | s_{1,j}, s_{2,j}, x_{2,j}, x_{3,j}, w_{3,j}) p(\tilde{x}_{1,j} | s_{1,j}, s_{2,j}, x_{2,j}, x_{3,j}, w_{3,j}) p(s_{1,j}, s_{2,j}, x_{2,j}, x_{3,j}, w_{3,j}). \quad (\text{D.4})$$

Using the chain (D.4) in [35, Thm. 14.2.3], $\Pr \{E_2^r(\tilde{u}_1) | (E_1^r)^c\}$ can be bounded by

$$\Pr \{E_2^r(\tilde{u}_1) | (E_1^r)^c\} \leq 2^{-n[I(X_1; Y_3 | S_1, S_2, X_2, X_3, W_3) - \epsilon_0]} \\ \stackrel{(a)}{=} 2^{-n[I(X_1; Y_3 | S_1, X_2, X_3) - \epsilon_0]}, \quad (\text{D.5})$$

where (a) follows from the Markov chain $(S_2, W_3) - (S_1, X_2, X_3) - Y_3$. Hence,

$$\Pr \{E_2^r | (E_1^r)^c\} \leq \sum_{\tilde{u}_1 \in \{1, 2, \dots, 2^{nR_1}\}, \tilde{u}_1 \neq u_{1,b}} \Pr \{E_2^r(\tilde{u}_1) | (E_1^r)^c\} \\ \leq 2^{n[R_1 - (I(X_1; Y_3 | S_1, X_2, X_3) - \epsilon_0)]}, \quad (\text{D.6})$$

which can be bounded by ϵ , for large enough n , if

$$R_1 < I(X_1; Y_3 | S_1, X_2, X_3) - \epsilon_0. \quad (\text{D.7})$$

Following similar arguments as in (D.2)–(D.7), it follows that $\Pr \{E_3^r | (E_1^r)^c\}$ can be bounded by ϵ , for large enough n , if

$$R_2 < I(X_2; Y_3 | S_2, X_1, X_3) - \epsilon_0, \quad (\text{D.8})$$

and $\Pr \{E_4^r | (E_1^r)^c\}$ can be bounded by ϵ , for large enough n , if

$$R_1 + R_2 < I(X_1, X_2; Y_3 | S_1, S_2, X_3) - \epsilon_0. \quad (\text{D.9})$$

We conclude that when conditions (D.7), (D.8) and (D.9) hold, for large enough n , $\bar{P}_{r, ch}^{(n)} \leq 4\epsilon$.

Relay source decoder: The relay source decoder error probability analysis is similar to the error probability analysis of the relay source decoder in the separation-based achievable rate detailed in Section III. The average probability of source decoder error at the relay can be made arbitrarily small, for sufficiently large n , if

$$H(S_1 | S_2, W_3) + \epsilon_0 < R_1, \quad (\text{D.10a})$$

$$H(S_2 | S_1, W_3) + \epsilon_0 < R_2, \quad (\text{D.10b})$$

$$H(S_1, S_2 | W_3) + \epsilon_0 < R_1 + R_2. \quad (\text{D.10c})$$

Combining conditions (D.7), (D.8) and (D.9) with conditions (D.10) yields the relay decoding constraints (42a)–(42c) in Thm. 8, for the achievability of rate $\kappa = 1$.

Next the destination error probability analysis is derived.

Destination error probability: The average decoding error probability at the destination in block b , $\bar{P}_d^{(n)}$, is defined by

$$\begin{aligned} \bar{P}_d^{(n)} &= \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \in \mathcal{S}_1^n \times \mathcal{S}_2^n} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \Pr \{ \text{destination error} | (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \text{ are the source outputs} \} \\ &\leq 2\epsilon + \sum_{(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_b) \in A_\epsilon^{*(n)}(S_1, S_2, W)} p(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_b) \times \\ &\quad \Pr \{ \text{destination error} | (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) \text{ are the source outputs, } \mathbf{w}_b \text{ is the side information} \}, \end{aligned} \quad (\text{D.11})$$

where (D.11) follows arguments similar to those led to (C.1d).

In the following we show that for $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_b) \in A_\epsilon^{*(n)}(S_1, S_2, W)$, the summation terms in (D.11) can be upper bounded independently of $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_b)$. In order to show the above we assume that $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_b) \in A_\epsilon^{*(n)}(S_1, S_2, W)$ and let \mathcal{D} denote the event that the triplet $(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{w}_b)$ is the sources outputs and the side information at the destination.

Let ϵ_0 be a positive number such that $\epsilon_0 > \epsilon$ and $\epsilon_0 \rightarrow 0$ as $\epsilon \rightarrow 0$. Assuming correct decoding at block $b+1$ (hence $(u_{1,b+1}, u_{2,b+1})$ are available at the destination), a destination decoding error event, E^d , in block b , is the union of the following events:

$$\begin{aligned} E_1^d &\triangleq \{ (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}, \mathbf{x}_1(u_{1,b+1}, \mathbf{s}_{1,b}), \mathbf{x}_2(u_{2,b+1}, \mathbf{s}_{2,b}), \mathbf{x}_3(\mathbf{s}_{1,b}, \mathbf{s}_{2,b}), \mathbf{w}_b, \mathbf{y}_{b+1}) \notin A_\epsilon^{*(n)} \}, \\ E_2^d &\triangleq \{ (\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \neq (\mathbf{s}_{1,b}, \mathbf{s}_{2,b}) : (\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_1(u_{1,b+1}, \tilde{\mathbf{s}}_1), \mathbf{x}_2(u_{2,b+1}, \tilde{\mathbf{s}}_2), \mathbf{x}_3(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2), \mathbf{w}_b, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)} \}. \end{aligned}$$

From the AEP, for sufficiently large n , $\Pr \{ E_1^d | \mathcal{D} \}$ can be bounded by ϵ . Therefore by the union bound

$$\Pr \{ E^d | \mathcal{D} \} \leq \epsilon + \Pr \{ E_2^d | \mathcal{D}, (E_1^d)^c \}. \quad (\text{D.12})$$

The event E_2^d is the union of the following events:

$$\begin{aligned} E_{21}^d &\triangleq \{ \exists \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b} : (\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b}, \mathbf{x}_1(u_{1,b+1}, \tilde{\mathbf{s}}_1), \mathbf{x}_2(u_{2,b+1}, \mathbf{s}_{2,b}), \mathbf{x}_3(\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b}), \mathbf{w}_b, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)} \}, \\ E_{22}^d &\triangleq \{ \exists \tilde{\mathbf{s}}_2 \neq \mathbf{s}_{2,b} : (\mathbf{s}_{1,b}, \tilde{\mathbf{s}}_2, \mathbf{x}_1(u_{1,b+1}, \mathbf{s}_{1,b}), \mathbf{x}_2(u_{2,b+1}, \tilde{\mathbf{s}}_2), \mathbf{x}_3(\mathbf{s}_{1,b}, \tilde{\mathbf{s}}_2), \mathbf{w}_b, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)} \}, \\ E_{23}^d &\triangleq \{ \exists \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b}, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_{2,b} : (\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{x}_1(u_{1,b+1}, \tilde{\mathbf{s}}_1), \mathbf{x}_2(u_{2,b+1}, \tilde{\mathbf{s}}_2), \mathbf{x}_3(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2), \mathbf{w}_b, \mathbf{y}_{b+1}) \in A_\epsilon^{*(n)} \}. \end{aligned}$$

Hence, by the union bound

$$\Pr \{ E_2^d | \mathcal{D}, (E_1^d)^c \} \leq \sum_{j=1}^3 \Pr \{ E_{2j}^d | \mathcal{D}, (E_1^d)^c \}. \quad (\text{D.13})$$

To bound $\Pr \{ E_{21}^d | \mathcal{D}, (E_1^d)^c \}$ we define the event $E_{21}^d(\tilde{\mathbf{s}}_1)$ as follows

$$\begin{aligned} E_{21}^d(\tilde{\mathbf{s}}_1) &\triangleq \{ (\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b}, \mathbf{x}_1(u_{1,b+1}, \tilde{\mathbf{s}}_1), \mathbf{x}_2(u_{2,b+1}, \mathbf{s}_{2,b}), \mathbf{x}_3(\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b}), \mathbf{w}_b, \mathbf{y}_{b+1}) \\ &\quad \in A_\epsilon^{*(n)}(S_1, S_2, X_1, X_2, X_3, W, Y) | \tilde{\mathbf{s}}_1 \neq \mathbf{s}_{1,b} \}. \end{aligned} \quad (\text{D.14})$$

Let us denote $\mathbf{s}_{2,b}$ by \mathbf{s}_2 , $\mathbf{x}_1(u_{1,b+1}, \tilde{\mathbf{s}}_1)$ by $\tilde{\mathbf{x}}_1$, $\mathbf{x}_2(u_{2,b+1}, \mathbf{s}_{2,b})$ by \mathbf{x}_2 , $\mathbf{x}_3(\tilde{\mathbf{s}}_1, \mathbf{s}_{2,b})$ by $\tilde{\mathbf{x}}_3$, \mathbf{w}_b by \mathbf{w} and \mathbf{y}_{b+1} by \mathbf{y} . Note that the joint distribution of $\tilde{\mathbf{s}}_1$, \mathbf{s}_2 , $\tilde{\mathbf{x}}_1$, \mathbf{x}_2 , $\tilde{\mathbf{x}}_3$, \mathbf{w} and \mathbf{y} obeys

$$p(\tilde{\mathbf{s}}_1, \mathbf{s}_2, \tilde{\mathbf{x}}_1, \mathbf{x}_2, \tilde{\mathbf{x}}_3, \mathbf{w}, \mathbf{y}) = \prod_{j=1}^n p(\tilde{s}_{1,j}, s_{2,j}, w_j) p(x_{2,j}, y_j | s_{2,j}, w_j) p(\tilde{x}_{1,j}, \tilde{x}_{3,j} | \tilde{s}_{1,j}, s_{2,j}, w_j). \quad (\text{D.15})$$

Equation (D.15) shows that the assumption of the Lemma [9, Lemma, Appendix A]³ on the conditional distribution is satisfied. Hence, we can use this lemma to bound $\Pr \{E_{21}^d(\tilde{\mathbf{s}}_1) | \mathcal{D}, (E_1^d)^c\}$ with the following assignments:

$$\mathbf{z}_1 = (\mathbf{s}_2, \mathbf{w}), \mathbf{z}_2 = \tilde{\mathbf{s}}_1, \mathbf{Z}_3 = \phi, \mathbf{Z}_4 = (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_3), \mathbf{Z}_5 = (\mathbf{X}_2, \mathbf{Y}), \quad (\text{D.16})$$

as follows

$$\Pr \{E_{21}^d(\tilde{\mathbf{s}}_1) | \mathcal{D}, (E_1^d)^c\} \leq 2^{-n[I(X_1, X_3; Y | S_2, X_2, W) - \epsilon_0]}. \quad (\text{D.17})$$

Hence

$$\begin{aligned} \Pr \{E_{21}^d | \mathcal{D}, (E_1^d)^c\} &\leq \sum_{\substack{\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1 \\ \tilde{\mathbf{s}}_1 \in A_\epsilon^{*(n)}(S_1 | \mathbf{s}_{2,b}, \mathbf{w}_b)}} \Pr \{E_{21}^d(\tilde{\mathbf{s}}_1) | \mathcal{D}, (E_1^d)^c\} \\ &\leq \sum_{\substack{\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1 \\ \tilde{\mathbf{s}}_1 \in A_\epsilon^{*(n)}(S_1 | \mathbf{s}_{2,b}, \mathbf{w}_b)}} 2^{-n[I(X_1, X_3; Y | S_2, X_2, W) - \epsilon_0]} \\ &\leq \left| A_\epsilon^{*(n)}(S_1 | \mathbf{s}_{2,b}, \mathbf{w}_b) \right| \cdot 2^{-n[I(X_1, X_3; Y | S_2, X_2, W) - \epsilon_0]} \\ &\leq 2^{n[H(S_1 | S_2, W) - I(X_1, X_3; Y | S_2, X_2, W) + 2\epsilon_0]}, \end{aligned} \quad (\text{D.18})$$

which can be bounded by ϵ , for large enough n , if

$$H(S_1 | S_2, W) < I(X_1, X_3; Y | S_2, X_2, W) - 2\epsilon_0, \quad (\text{D.19})$$

Following arguments similar to those that led to (D.18), we can also show that $\Pr \{E_{22}^d | \mathcal{D}, (E_1^d)^c\}$ can be bounded by ϵ , for large enough n , if

$$H(S_2 | S_1, W) < I(X_2, X_3; Y | S_1, X_1, W) - 2\epsilon_0, \quad (\text{D.20})$$

and $\Pr \{E_{23}^d | \mathcal{D}, (E_1^d)^c\}$ can be bounded by ϵ , for large enough n , if

$$H(S_1, S_2 | W) < I(X_1, X_2, X_3; Y | W) - 2\epsilon_0, \quad (\text{D.21})$$

Hence, if conditions (42a)–(42c) hold, for large enough n ,

$$\Pr \{E_2^d | \mathcal{D}, (E_1^d)^c\} \leq \sum_{j=1}^3 \Pr \{E_{2j}^d | \mathcal{D}, (E_1^d)^c\} \leq 3\epsilon. \quad (\text{D.22})$$

³The following Lemma is a slight variation of [9, Lemma, Appendix A]. *Lemma:* Let $(Z_1, Z_2, Z_3, Z_4, Z_5)$ be random variables with joint distribution $p(z_1, z_2, z_3, z_4, z_5)$. Fix $(\mathbf{z}_1, \mathbf{z}_2) \in A_\epsilon^{*(n)}(Z_1, Z_2)$, and let $\mathbf{Z}_3, \mathbf{Z}_4, \mathbf{Z}_5$ be drawn according to $\Pr \{\mathbf{Z}_3 = \mathbf{z}_3, \mathbf{Z}_4 = \mathbf{z}_4, \mathbf{Z}_5 = \mathbf{z}_5 | \mathbf{z}_1, \mathbf{z}_2\} = \prod_{j=1}^n p(z_{3,j} | z_{2,j}, z_{1,j}) p(z_{4,j} | z_{2,j}, z_{1,j}) p(z_{5,j} | z_{3,j}, z_{1,j})$. Then $\Pr \{(\mathbf{z}_1, \mathbf{z}_2, \mathbf{Z}_3, \mathbf{Z}_4, \mathbf{Z}_5) \in A_\epsilon^{*(n)}(Z_1, Z_2, Z_3, Z_4, Z_5)\} \leq 2^{-n[I(Z_3; Z_4 | Z_1, Z_2) + I(Z_5; Z_2, Z_4 | Z_1, Z_3) - 8\epsilon]}$. The proof follows arguments similar to the proof of [9, Lemma Appendix A].

Combining equations (D.11), (D.12) and (D.22) yields

$$\bar{P}_d^{(n)} \leq \Pr \{E_2^d | \mathcal{D}, (E_1^d)^c\} + 3\epsilon \leq 6\epsilon. \quad (\text{D.23})$$

REFERENCES

- [1] G. Kramer and A. J. Wijnngaarden, "On the white Gaussian multiple-access relay channel". *Proc. IEEE Int. Symp. Inform. Theory*, Sorrento, Italy, Jun. 2000, p. 40.
- [2] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks". *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [3] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Offset encoding for multiaccess relay channels". *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3814–3821, Oct. 2007.
- [4] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Capacity theorems for the multiple-access relay channel". *Proc. 42nd Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sep. 2004, pp. 1782–1791.
- [5] R. Tandon and H. Vincent Poor, "On the capacity region of multiple-access relay channels". *Proc. Conf. Information Sciences and Systems*, Baltimore, MD, March 2011.
- [6] C. E. Shannon, "A mathematical theory of communication". *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and pp. 623–656, 1948.
- [7] S. Shamai and S. Verdú, "Capacity of channels with side information". *European Transactions on Telecommunications and Related Technologies*, vol. 6, no. 5, pp. 587–600, Sep. 1995.
- [8] C. E. Shannon, "Two-way communication channels". *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, vol. 1, pp. 611–644, 1961.
- [9] T. M. Cover, A. El. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources". *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.
- [10] D. Gündüz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels". *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 3927–3944, Sep. 2009.
- [11] L. Sankar, N. B. Mandayam, and H. V. Poor, "On the sum-capacity of the degraded Gaussian multiaccess relay channel". *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5394–5411, Dec. 2009.
- [12] G. Dueck, "A note on the multiple access channel with correlated sources". *IEEE Trans. Inform. Theory*, vol. 27, no. 2, pp. 232–235, Mar. 1981.
- [13] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple access channel and related problems in multiuser information theory". *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 396–412, May. 1983.
- [14] T. S. Han and M. H. M. Costa, "Broadcast channels with arbitrarily correlated sources". *IEEE Trans. Inform. Theory*, vol. 33, no. 5, pp. 641–650, Sep. 1987.
- [15] G. Kramer and C. Nair, "Comments on broadcast channels with arbitrarily correlated sources". *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, July. 2009.
- [16] E. Tuncel, "Slepian-Wolf coding over broadcast channels". *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1469–1482, April 2006.
- [17] M. Salehi and E. Kurtas, "Interference channels with correlated sources". *Proc. IEEE Int. Symp. Inform. Theory*, San Antonio, TX, Jan. 1993.
- [18] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel". *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp.49–60, Jan, 1981.
- [19] W. Liu and B. Chen, "Interference channels with arbitrarily correlated sources". *Submitted to the IEEE Trans. Information Theory*, Nov. 2009.
- [20] N. Liu, D. Gündüz, A. Goldsmith and H. V. Poor, "Interference channels with correlated receiver side information". *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 5984–5998, Dec. 2010.
- [21] B. Smith and S. Vishwanath, "Capacity analysis of the relay channel with correlated sources". *Submitted to IEEE Trans. Inform. Theory*, Jan., 2007.
- [22] D. Gündüz and E. Erkip, "Reliable cooperative source transmission with side information". *Proc. IEEE Information Theory Workshop*, Bergen, Norway, Jul. 2007.

- [23] D. Gündüz, E. Erkip, A. Goldsmith and H. V. Poor, “Cooperative relaying in sensor networks”. *Proc. 5th Int. Conf. on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2010)*, Cannes, France, 2010.
- [24] D. Gündüz and E. Erkip, “Joint source–channel codes for MIMO block-fading channels”. *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 116–134, Jan. 2008.
- [25] R. Kwak, W. Lee, A. El Gamal, and J. Cioffi, “Relay with side information”. *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 606–610.
- [26] M. Sefidgaran, B. Akhbari, Y. Mohsenzadeh and M. R. Aref, “Reliable source transmission over relay networks with side information”. *Proc. IEEE Int’l Symp. Inform. Theory*, Seoul, South Korea, July 2009.
- [27] S. Wu and Y. Bar-Ness. “OFDM systems in the presence of phase noise: consequences and solutions”. *IEEE Trans. Commun.* vol. 52, no. 11, pp. 1988–1996 Nov. 2004.
- [28] U. Erez, M. D. Trott, and G. W. Wornell, “Rateless coding and perfect rate-compatible codes for Gaussian channels”. *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Jul. 2006, Seattle, WA, pp. 528–532.
- [29] B. Sklar, “Rayleigh fading channels in mobile digital communication systems part I: Characterization”. *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, Jul. 1997.
- [30] R. Dabora. “The capacity region of the fading interference channel with a relay in the strong interference regime”. Submitted to the *IEEE Trans. Information Theory*, June 2010. Available at <http://www.bgu.ac.il/~daborona>.
- [31] C. T. K. Ng, N. Jindal, A. J. Goldsmith, and U. Mitra, “Capacity gain from two-transmitter and two-receiver cooperation”. *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3822–3827, Oct. 2007.
- [32] G. Farhadi and N. C. Beaulieu, “On the ergodic capacity of wireless relaying systems over Rayleigh fading channels”. *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4462–4467, Nov. 2008.
- [33] G. Farhadi and N. C. Beaulieu, “On the ergodic capacity of multi-hop wireless relaying systems”. *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2286–2291, May 2009.
- [34] I. Stanojev, O. Simeone, Y. Bar-Ness, and C. You, “Performance of multi-relay collaborative hybrid-ARQ protocols over fading channels”. *IEEE Communications Letters*, vol. 10, no. 7, pp. 522–524 Jul. 2006.
- [35] T. M. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons Inc., 1991.
- [36] R. W. Yeung, *A First Course in Information Theory*. Springer, 2002.
- [37] F. D. Nesser and J. L. Massey, “Proper complex random processes with applications to information theory”. *IEEE Trans. Inform. Theory*, vol 39, no. 4, pp. 1293–1302, Jul. 1993.
- [38] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources”. *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [39] A. El. Gamal and Y. H. Kim, “Lecture Notes on Network Information Theory”. arXiv ref. arXiv:1001.3404v3.
- [40] J. L. Massey, “Causality, feedback and directed information”. *Proceedings of the IEEE International Symposium on Information Theory and its Applications*, Nov. 1990, Waikiki, HI, pp. 303–305.
- [41] T. Berger, “Multiterminal source coding”. in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1977.
- [42] Y. Oohama, “Gaussian multiterminal source coding”. *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.
- [43] M. Abramowitz and I. A. Stegun, (Eds.) *Exponential Integral and Related Functions*. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, 1972.